Psychometrika

CONTENTS

Continued on next page

VOLUME TWENTY-FOUR SEPTEMBER 1959 NUMBER 3

Psychometrika

CONTENTS (Cont.)

VOLUME TWENTY-FOUR SEPTEMBER 1959 NUMBER 3

PSYCHOMETRIKA—VOL. 24, NO. 3
SEPTEMBER, 1959

ITEM SELECTION PROCEDURES FOR ITEM VARIABLES WITH A KNOWN FACTOR STRUCTURE*

G. ELFVING
UNIVERSITY OF HELSINKI

R. SITGREAVES AND H. SOLOMON
COLUMBIA UNIVERSITY

This paper discusses the item selection problem when the item responses follow a linear multiple factor model. Because of this restrictive assumption, not too unrealistic in situations such as mental testing, it is possible to select optimal sets of items without going through all possible combinations. A method proposed by Elfving to accomplish this is analyzed and then demonstrated through the use of two illustrations. The common and often used procedure of observing the magnitude of the correlation coefficient as an index in item selection is shown to have some merit in the single-factor case.

The question of type and number of items to be used in a test of mental ability, an attitude scale, a personality inventory, or biographical inventory is a familiar topic in psychological testing. This paper will attempt some resolution of the problem for the restricted situation where the item variables obey a known factor structure. For simplicity the generic term *test* will be taken to represent a collection of items whether dealing with attitude, personality, etc. The classical situation where test response is a linear function of the item responses will be considered. This, of course, may be somewhat unrealistic in some situations, e.g., biographical inventories may call for nonclassical approaches in item selection. On the other hand, even for what may be termed the classical situation the question has not been resolved.

Assume that the test response is to be used as a predictor of a criterion variable. The number and type of items selected for a test, therefore, are governed by the usefulness of the resulting test response for prediction purposes. In the usual techniques of item selection, the size of a correlation coefficient between an item response and the criterion indicates how well the item aids in the prediction made by the test. While the correlation coefficient appears to have pragmatic value for tests of mental ability, it falls short in a number of other testing situations. It should be remembered that the correlation coefficient is a methodological tool borrowed from those who conceptualized its use in anthropometric settings; it may not be efficient in all psychometric situations. A conceptualization of the problem which

*Work performed under contract AF 41(657)-244 with the School of Aviation Medicine, Randolph AFB, Texas.

leads to a more general approach to item selection will be considered here.

Suppose for prediction of a criterion z, a large number N of nonrepeatable observations item variables x_i is potentially available. The quantities $x_1 , x_2 , \cdots , x_N ; z$ are assumed to be random variables having joint distributions which are known except perhaps for certain parameters, from past experience or by assumption. In other words, sampling questions are completely disregarded, and we restrict ourselves to a design problem only.

For practical reasons, one wants to base the prediction of z on a restricted number, say $n < N$, of observations x_i . The problem is how to choose them. In a psychological application one can think of the x_i as scores obtained as responses to items on a reading test, and the criterion z as the school grade in which a child will be classified on the basis of a moderate number of items. For concreteness, one can think of N as being of the order 100 to 1000, and n from 10 to 50.

The proposed item selection approach is based on the assumption that the variables $x_1 , x_2 , \cdots , x_N ; z$ have a known factor structure, with a comparatively small number k of common factors; assume that, practically speaking, k ranges from 1 to 5. The assumption of this latent factor universe, which is a restrictive assumption, and how this added information provides for item selection is a departure from other item selection methods. As will be seen, it also provides a rationale in one situation for the correlation coefficient, used so extensively at present. This paper is essentially an analytical and expository account of a method proposed by Elfving [1, 2, 5, 6] and is based on some prior work in different contexts [3, 4].

In the present paper, an explanation of the principles leading to the steps of the method is given, and the method is applied to two sets of data. One set comes from the Educational Testing Service in connection with aptitude testing for law school, where the best two out of six items are to be selected. The other set was artificially constructed; from ten items the best four, the best five, and the best six are, respectively, to be selected. It can easily be demonstrated [1] that the best set of $(n + 1)$ items need not contain the best set of n items. One reason for the scarcity of data on which to employ our item selection procedure is that not only must the factor structure for N items be known completely but the factor loadings for the criterion variable z must be known or guessed realistically. The procedure to be described is also useful for selecting tests from a battery of tests and in these situations, complete factor structures are available. However the language and demonstrations will stress the selection of items for a test.

The Factor Structure and Prediction Criterion

Consider the following factor structure for the x_i and z. Naturally if the items are to have any validity in predicting the criterion, they should be

composed of the same common factors and differ at most in factor loadings and specific factors. Thus,

$$x_i = a_{i1}y_1 + \cdots + a_{ik}y_k + \epsilon_i \qquad (i = 1, \cdots, N), \tag{1}$$

$$z = c_1y_1 + \cdots + c_ky_k + \eta, \tag{2}$$

where

(*i*) the *loadings* a_{ij} , c_j are known constants;

(*ii*) the (unobservable) *specific factors* $\epsilon_1, \cdots, \epsilon_N, \eta$ are random variables with mean zero, distributed independently of the unobservable (latent) common factors $y_1, \cdots, y_k$;

(*iii*) the ϵ_i have known covariance matrix (in most of what follows, this will be assumed to be diagonal);

(*iv*) η has variance σ_e^2 and is uncorrelated with the ϵ_i .

For the common factors y_j , consider two different models which, however, lead essentially to the same selection technique;

(*v* a) fixed constants model where the y_j are unknown constants;

(*v* b) random factor model where the y_j are random variables, with mean zero and a known nonsingular covariance matrix T.

In model (*v* a) the y_j may be thought of as factor values pertaining to the particular individual for which z has to be predicted. In (*v* b) the individual is thought of as belonging to a population with known characteristics.

In predicting z, only linear unbiased minimum-variance predictors will be considered; i.e., for any selected set of n items out of N items, a predictor, is taken to be the linear combination

$$\hat{z} = \sum_{i=1}^{n} q_i x_i \ , \tag{3}$$

which satisfies both $E(\hat{z} - z) = 0$, and $E(\hat{z} - z)^2 =$ minimum. This is a reasonable criterion for estimation quite often used in classical multivariate analysis. Suppose for the selected set of n items in matrix notation,

$$x = Ay + \epsilon, \tag{4}$$

$$z = c'y + \eta, \tag{5}$$

and

$$\Sigma = \operatorname{cov} \epsilon = E(\epsilon\epsilon'). \tag{6}$$

Then, in the fixed constants model, the best predictor of z on the basis of x is

$$\hat{z} = c'(A'\Sigma^{-1}A)^{-1}A'\Sigma^{-1}x, \tag{7}$$

with prediction variance

$$E(\hat{z} - z)^2 = c'(A'\Sigma^{-1}A)^{-1}c + \sigma_e^2 \ . \tag{8}$$

In the random factor model, the best predictor of z, on the basis of x, is

$$\hat{z} = c'(T^{-1} + A'\Sigma^{-1}A)^{-1}A'\Sigma^{-1}x, \tag{9}$$

with prediction variance

$$E(\hat{z} - z)^2 = c'(T^{-1} + A'\Sigma^{-1}A)^{-1}c + \sigma_e^2 . \tag{10}$$

These results are merely stated since they are well known in multivariate analysis.

The Selection Region

Now consider the selection problem. The problem is to choose n out of N items. There are many possibilities but that set of n items is desired which will minimize the prediction variance $E(\hat{z} - z)^2$. Naturally a procedure is preferred which makes it unnecessary to look at all possible sets of n items to arrive at the proper choice. Considerations of all possibilities can be done for small n and small N, but even when $N = 10$ and $n = 2$, it is quite tedious.

The prediction variance is given, for the two models considered, by (8) and (10). Since σ_e^2 does not depend on the selection, the problem reduces to the minimization of

$$V = c'M^{-1}c, \tag{11}$$

where

$$M = M_0 + A'\Sigma^{-1}A, \tag{12}$$

with $M_0 = 0$ in the parametric model, and $M_0 = T^{-1}$ in the random factor model. The variable elements are, of course, in the $n \times k$ matrix A, and the $n \times n$ matrix Σ, both of which depend on the choice of items.

From now on, assume that the specific factors are uncorrelated, with variances σ_i^2 . That is, for a selected set of n items,

$$A'\Sigma^{-1}A = \begin{bmatrix} \sum_{i=1}^{n} \frac{a_{i1}^2}{\sigma_i^2} & \sum_{i=1}^{n} \frac{a_{i1}a_{i2}}{\sigma_i^2} & \cdots & \sum_{i=1}^{n} \frac{a_{i1}a_{ik}}{\sigma_i^2} \\ \sum_{i=1}^{n} \frac{a_{i1}a_{i2}}{\sigma_i^2} & \sum_{i=1}^{n} \frac{a_{i2}^2}{\sigma_i^2} & \cdots & \sum_{i=1}^{n} \frac{a_{i2}a_{ik}}{\sigma_i^2} \\ \cdots & \cdots & \cdots & \cdots \\ \sum_{i=1}^{n} \frac{a_{i1}a_{ik}}{\sigma_i^2} & \sum_{i=1}^{n} \frac{a_{i2}a_{ik}}{\sigma_i^2} & \cdots & \sum_{i=1}^{n} \frac{a_{ik}^2}{\sigma_i^2} \end{bmatrix}. \tag{13}$$

Denote by a_i row i of A written as a column vector. Then,

$$A'\Sigma^{-1}A = \sum_{i=1}^{n} \frac{a_i a_i'}{\sigma_i^2} = \sum_{i=1}^{n} u_i u_i' , \quad \text{where} \quad u_i = \frac{a_i}{\sigma_i} ; \tag{14}$$

that is,

$$u_i u_i' = \begin{bmatrix} \frac{a_{i1}^2}{\sigma_i^2} & \frac{a_{i1}a_{i2}}{\sigma_i^2} & \cdots & \frac{a_{i1}a_{ik}}{\sigma_i^2} \\ \frac{a_{i1}a_{i2}}{\sigma_i^2} & \frac{a_{i2}^2}{\sigma_i^2} & \cdots & \frac{a_{i2}a_{ik}}{\sigma_i^2} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{a_{i1}a_{ik}}{\sigma_i^2} & \frac{a_{i2}a_{ik}}{\sigma_i^2} & \cdots & \frac{a_{ik}^2}{\sigma_i^2} \end{bmatrix} \tag{15}$$

is, for every i, a $k \times k$ matrix of rank 1. The sum $\sum_{i=1}^{n} u_i u_i'$ has for elements the moments of the k-dimensional item population consisting of the points $u_1, \cdots u_n$, each with weight 1.

Returning to the original numbering of the items, $i = 1, \cdots, N$, and denoting by ω the selected set of n subscripts, (12) may now be written

$$M = M_0 + \sum_{\omega} u_i u_i' . \tag{16}$$

The $k \times k$ matrix M is the *information matrix* of the experiment, i.e., of the selected set of observations. The sum $\Sigma u_i u_i'$ represents the information offered by the items in ω. In the random factor model, $M_0 = T^{-1}$ may be said to represent the a priori information contained in the assumption that the factors are random variables with zero means and covariance matrix T. It may be noted, incidentally, that the constant term M_0 in (16) can also be used to take care of any *fixed* source of information, such as possible "compulsory" items.

It is seen that the effect of a particular item-number i, say, depends solely on the vector u_i, i.e., on the reduced loading vector obtained by standardizing the item variable to have variance equal to one. In the fixed constants case, one may multiply the u_i by a common factor without affecting the minimization problem at hand. As a consequence, the item variables may actually be standardized to any common variance, not necessarily unity.

Also note that two items, one with vector u_i and the other with vector $-u_i$, yield the same contribution to the information matrix. For this reason and reasons of symmetry which will become clearer later on, each item will usually be described by means of the *pair* of opposite points $\pm u_i$. These points, in k-space, will be referred to as *item points*.

It is natural to think of the selected set ω of item points as occupying a certain "selection region" S which will, of course, depend in some way on the totality of available item points. Before attacking the question of how to find this region, it may be useful to discuss briefly the cases $k = 1$ and $k = 2$.

When $k = 1$, all symbols in (11) and (16) denote scalars, and the problem reduces to minimizing

$$V = \frac{c^2}{m_0 + \sum_{\omega} u_i^2} \tag{17}$$

by a proper choice of ω. Thus one has simply to select the n items with largest $| u_i |$. If the x_i and u_i, as often in factor analysis, are standardized to variance one, $a_i^2 + \sigma_i^2 = 1$, and hence $u_i^2 = a_i^2/\sigma_i^2 = a_i^2/(1 - a_i^2)$. The items with largest $| u_i |$ are then the same as those with largest $| a_i |$, i.e., those having largest loadings with respect to the single common factor. In this case, however,

$$\frac{a_i^2}{1 - a_i^2} = \frac{\rho_{x_i y_1}^2}{1 - \rho_{x_i y_1}^2}, \tag{18}$$

so that this procedure is equivalent to picking the x_i having the highest absolute correlation with y_1. This procedure is usually followed by psychologists, except that the latent factor is not available. Therefore, resort is made to a manifest equivalent, usually observed total test score.

In the case of $k = 2$, it is clear, in general, that if an item point happens to lie precisely on the straight line determined by the vector c, say $u_1 = \lambda_1 c$, then

$$\begin{aligned} x_1 &= a_{11}y_1 + a_{12}y_2 + \epsilon_1 \\ &= \sigma_1\lambda_1(c_1y_1 + c_2y_2) + \epsilon_1 , \end{aligned} \tag{19}$$

and $x_1/\sigma_1\lambda_1$ provides an unbiased estimate of $c'y$. The variance of this estimate is $\sigma_1^2/(\sigma_1\lambda_1)^2 = 1/\lambda_1^2$. Accordingly, item points along the line c may be expected to contribute more to the estimation of $c'y$, and hence to the prediction of z, the farther out they are along the line.

Similarly, suppose there are two item points located more or less symmetrically with respect to the line c. For example, suppose

$$\begin{aligned} u_1 &= \left(\frac{\lambda_{c_1}}{\sqrt{c_1^2 + c_2^2}} + kc_2 \qquad \frac{\lambda_{c_2}}{\sqrt{c_1^2 + c_2^2}} - kc_1\right), \\ u_2 &= \left(\frac{\lambda_{c_1}}{\sqrt{c_1^2 + c_2^2}} + kc_2 \qquad \frac{\lambda_{c_2}}{\sqrt{c_1^2 + c_2^2}} + kc_1\right). \end{aligned} \tag{20}$$

Then

$$\frac{1}{2}\left(\frac{x_1}{\sigma_1} + \frac{x_2}{\sigma_2}\right) = \frac{\lambda}{\sqrt{c_1^2 + c_2^2}}(c_1y_1 + c_2y_2) + \frac{1}{2}\left(\frac{\epsilon_1}{\sigma_1} + \frac{\epsilon_2}{\sigma_1}\right), \tag{21}$$

and

$$\frac{\sqrt{c_1^2 + c_2^2}}{2\lambda}\left(\frac{x_1}{\sigma_1} + \frac{x_2}{\sigma_2}\right) \tag{22}$$

is an unbiased estimate of $c'y$ with variance

$$\frac{(c_1^2 + c_2^2)}{2\lambda^2}. \tag{23}$$

Since λ is the distance from the origin to the point of intersection of the c-vector and the line joining u_1 and u_2, it is clear again that the two item points, after the elimination of the orthogonal component, provide an estimator of $c'y$ which is better the farther off in the $\pm$ c-direction they lie.

These heuristic remarks make it plausible that the selection region S will have to comprise the outer parts of k-space with regard to the directions $\pm$ c and will have to fulfill the additional requirement that the item points in S should in some way balance each other with respect to that direction. It turns out [1, 2, 5, 6] that S may be taken to consist of two symmetrical half-spaces (a twin half-space) bounded by two parallel planes $t'u = \pm h$; the direction t of their common normal will depend in a certain manner on the item points included in the selection region, the latter becoming in this way implicitly determined.

In order to formulate the result just sketched, it is necessary to introduce a continuation device which will lead to a simple optimizing criterion. This will greatly facilitate the solution of our problem without essentially affecting the practical application. For this purpose consider, instead of the previous M defined by (16), the generalized information matrix

$$M = M_0 + \sum_{i=1}^{N} p_i u_i u_i', \tag{24}$$

where the allocation vector $p = (p_1 \cdots p_N)$ is subject to the restrictions

$$0 \leq p_i \leq 1, \qquad \sum_{1}^{N} p_i = n. \tag{25}$$

Obviously, the set of matrices given by (24) and (25) contains the set of matrices (16). For any particular p, those items for which $p_i = 1$ will be referred to as *totally selected*, those for which $p_i = 0$, as *nonselected*, and those for which $0 < p_i < 1$, as *fractionally selected*.

The following interpretation of the fractional p_i (in the parametric case) may be instructive. Imagine for a moment that the observations x_i may be independently repeated, each of them at most r times, and assume that a total of rn observations is allowed. Let $n_1, \cdots, n_N$ ($0 \leq n_i \leq r$; $\sum n_i = rn$) be the number of times that the different observations are repeated. The information matrix of the resulting experiment may be written

$$M = \Sigma n_i u_i u_i' = r\Sigma p_i u_i u_i', \tag{26}$$

where the $p_i = n_i/r$ vary from 0 to 1 through multiples of $1/r$, subject to the condition $\Sigma p_i = n$. Since the factor r in M is obviously irrelevant to the

minimization of (11), for large r, the problem is essentially a modified selection version of that formulated above. In particular, if $n = 1$, and r is large, one is concerned with the allocation problem treated by Elfving in [3]; the earlier problem thus appears a special case of the present one.

After these preparations, one may state the following propositions [2, 5].

THEOREM 1. *The scalar V, defined by*

$$V = c'M^{-1}c, \qquad \text{where} \quad M = M_0 + \sum_{i=1}^{N} p_i u_i u_i' \tag{27}$$

as a function of the vector p, has a minimum on the domain (25). In order for the allocation vector p to yield this minimum, it is necessary and sufficient that, for a certain number $h > 0$,

$$p_i = \begin{cases} 1 & \text{whenever} \quad | c'M^{-1}u_i | > h \\ 0 & \text{whenever} \quad | c'M^{-1}u_i | < h. \end{cases} \tag{28}$$

Moreover, there exists always a minimizing p with at most k fractional components.

Recall that M is a $k \times k$ matrix, hence $c'M^{-1}$ is a k-dimensional row vector, and $c'M^{-1}u_i$ is a linear form in the components of the vector u_i . The content of the theorem is that the selection region consists mainly of that part of k-space which lies outside two parallel hyperplanes $c'M^{-1}u = \pm h$; the item points lying in the boundary planes will have to be totally selected, fractionally selected, or nonselected, as the case may be. Since the fractional p_i (when they are taken to be as few as possible) can total at most k, there will be from $n - k + 1$ to n totally selected items. In practice, one may round off some or all of the fractional p_i to unity, i.e., select the corresponding items on an equal basis with the rest. In the latter case, at the expense of making at most $k - 1$ more observations than originally planned, one will be sure to achieve a variance V not exceeding the smallest one that could be attained by any total selection of n items. In those cases where the optimal p will contain only 1's and 0's, an exact solution of the original discrete problem is provided.

For a proof of Theorem 1 in the parametric case (the random factor case goes quite similarly) the reader is referred to [2, 5]. It should be noted that the theorem gives a necessary and sufficient criterion for optimum solutions, but no method for finding such a solution. When $k = 2$ and n is small, a graphic picture of the item points may lead to a good guess. For more complex situations, a method has been suggested [1, 6] based on the idea that the population of item points may be approximately described by a k-dimensional normal distribution with the same second-order moments. The method may be condensed into the following practical rule.

(*i*) Find the matrix Λ with elements

$$\lambda_{jh} = \sum_{i=1}^{N} u_{ij}u_{ih} \qquad (j, h = 1, \cdots, k). \tag{29}$$

(*ii*) Find the vector $\gamma = \Lambda^{-1}c$, i.e., solve the equations

$$\begin{aligned} \lambda_{11}\gamma_1 + \cdots + \lambda_{1k}\gamma_k &= c_1 , \\ &\cdots \\ \lambda_{k1}\gamma_1 + \cdots + \lambda_{kk}\gamma_k &= c_k . \end{aligned} \tag{30}$$

(*iii*) Find, for each i, the quantity

$$w_i = \gamma' u_i = \sum_j \gamma_j u_{ij} \tag{31}$$

and select the n items with largest $| w_i |$.

The selection found in this way should provide a good first guess and, of course, may be checked by means of Theorem 1. If the criterion is not fulfilled, one may try to improve the solution by exchanging one or more of the selected item points for others in the neighborhood of the boundary planes. An artificial example of such a procedure is given in a later section.

A Realistic Example when $k = 2$

To illustrate the meaning and use of Theorem 1, two examples will be considered, both of which involve a two-factor structure and assume a fixed constants model. The first example is based on data made available by the Educational Testing Service. These data resulted from responses to six items used to measure aptitude for success in law school and a criterion variable which measured success in law school. A factor analysis of the six items and the criterion variable resulted in two common factors plus specific factors. For this illustration,

$$\begin{aligned} x_i &= a_i' y + \epsilon_i \qquad (i = 1, 2, \cdots, 6), \\ z &= c' y + \eta, \end{aligned} \tag{32}$$

where

$$\begin{aligned} &a_1' = (.848 \quad .225), \quad a_2' = (.833 \quad .195), \quad a_3' = (.840 \quad .122), \\ &a_4' = (.481 \quad -.216), \quad a_5' = (.647 \quad -.172), \quad a_6' = (.869 \quad .204), \\ &c' = (.641 \quad -.205). \end{aligned} \tag{33}$$

The (unobservable) specific factors $\epsilon_1, \cdots, \epsilon_6$; η are assumed to be independently distributed random variables with mean zero. The variances of $\epsilon_1, \cdots, \epsilon_6$ are

$$\begin{aligned} &\sigma_1^2 = .230, \quad \sigma_2^2 = .268, \quad \sigma_3^2 = .280, \\ &\sigma_4^2 = .722, \quad \sigma_5^2 = .552, \quad \sigma_6^2 = .203. \end{aligned} \tag{34}$$

The vector of (unobservable) common factors $y' = (y_1 y_2)$ is assumed to be a vector of unknown constants.

Then, if $u_i = a_i/\sigma_i$,

$$(35)\quad \begin{aligned} u_1' &= (1.767 \quad .469), \quad u_2' = (1.608 \quad .376), \quad u_3' = (1.588 \quad .231), \\ u_4' &= (.566 \quad -.254), \quad u_5' = (.871 \quad -.231), \quad u_6' = (1.927 \quad .452); \end{aligned}$$

and

$$(36)\quad \begin{aligned} u_1u_1' &= \begin{bmatrix} 3.122 & .829 \\ .829 & .220 \end{bmatrix}, & u_2u_2' &= \begin{bmatrix} 2.586 & .605 \\ .605 & .141 \end{bmatrix}, \\ u_3u_3' &= \begin{bmatrix} 2.522 & .367 \\ .367 & .053 \end{bmatrix}, & u_4u_4' &= \begin{bmatrix} .320 & -.144 \\ -.144 & .065 \end{bmatrix}, \\ u_5u_5' &= \begin{bmatrix} .759 & -.201 \\ -.201 & .053 \end{bmatrix}, & u_6u_6' &= \begin{bmatrix} 3.713 & .871 \\ .871 & .204 \end{bmatrix}. \end{aligned}$$

The item points u_1 , $\cdots$, u_6 , and the vector c are shown graphically in Figure 1.

For a given set of x's, say x_{i_1} , $\cdots$, x_{i_n} , the best predictor of z on the basis of x is

$$(37)\quad z = c'M_\omega^{-1}U_\omega x^*,$$

where ω denotes the selected set of subscripts, and

$$M_\omega = \sum_\omega u_i u_i' ; \qquad U_\omega = (u_{i_1} \quad u_{i_2} \quad \cdots \quad u_{i_n}), \text{ a } 2 \times n \text{ matrix};$$

$$x^* = \left(\frac{x_{i_1}}{\sigma_{i_1}} \quad \frac{x_{i_2}}{\sigma_{i_2}} \quad \cdots \quad \frac{x_{i_n}}{\sigma_{i_n}}\right).$$

The variance of the estimate is

$$(38)\quad E(z - \hat{z})^2 = c'M^{-1}c + \sigma_e^2 .$$

M_ω may be considered the information matrix for the selected set of observations.

Consider the generalized information matrix

$$(39)\quad M_p = \sum_{i=1}^{6} p_i u_i u_i' ,$$

where the allocation vector $p' = (p_1 \quad p_2 \quad \cdots \quad p_6)$ is subject to the restrictions

$$(40)\quad 0 \leq p_i \leq 1, \qquad \sum_{i=1}^{6} p_i = 2.$$

That is, two out of the six items are desired. Then, Theorem 1 states that

$$(41)\quad V = c'M_p^{-1}c,$$

as a function of the vector p, has a minimum on the domain (40). In order for the allocation vector $p = p^*$ to yield this minimum, it is necessary and sufficient that for a certain number $h > 0$

$$p_i = \begin{cases} 1 & \text{whenever} \quad |c' M_{p^*}^{-1} u_i| > h \\ 0 & \text{whenever} \quad |c' M_{p^*}^{-1} u_i| < h, \end{cases}$$

where p^* is the minimizing vector. That is, the wholly selected item points lie outside of two parallel lines which are symmetric with respect to the origin, while the totally nonselected item points lie between these lines. (Totally selected and totally nonselected item points may also lie on the boundaries, i.e., parallel lines.) Any fractionally selected item, say n_i, must be on one of the lines defined by $| c' M_{p^*}^{-1} u_i | = h$. At most, two items may need to be fractionally selected.

Suppose in the present example, where $n = 2$, we first limit ourselves to picking the best two wholly selected items; that is, the possibility of fractionally selected items is not permitted. If one examines the 15 possible combinations of the two items in pairs, it is seen that items 4 and 5 are best, i.e., they yield the estimate with the smallest variance. Writing $p_0 = (0, 0, 0, 1, 1, 0)$, the variance of this estimate is

$$c' M_{p_0}^{-1} c = .381. \tag{42}$$

Theorem 1 states that if, in fact, p_0 is the minimizing vector when fractionally selected items are admitted, there is a number h such that

$$| c' M_{p_0}^{-1} u_4 | > h, \quad | c' M_{p_0}^{-1} u_5 | > h, \quad | c' M_{p_0}^{-1} u_j | < h \quad (j = 1, 2, 3, 6). \tag{43}$$

However,

$$\begin{aligned} &| c' M_{p_0}^{-1} u_1 | = 1.043, \quad | c' M_{p_0}^{-1} u_2 | = .950, \quad | c' M_{p_0}^{-1} u_3 | = .939, \\ &| c' M_{p_0}^{-1} u_4 | = .337, \quad | c' M_{p_0}^{-1} u_5 | = .517, \quad | c' M_{p_0}^{-1} u_6 | = 1.141. \end{aligned} \tag{44}$$

It follows that p_0 is not the minimizing vector, and it will be necessary to consider fractionally selected items.

Since items 4 and 5 are the best pair of items (the variance of the estimate based on items 2 and 6, for example, is 447.01), it seems reasonable that these items be included when fractional selection is allowed. The theorem indicates that there must be a pair of parallel lines such that the two fractionally selected items lie on the line.

It is seen in Figure 1 that if the two parallel lines are drawn through items 3 and 4, item 5 lies outside the lines and items 1, 2, and 6 lie between them. Thus one is led to attempt a fractional solution of the form

$$p^* = (0 \quad 0 \quad r \quad 1 - r \quad 1 \quad 0). \tag{45}$$

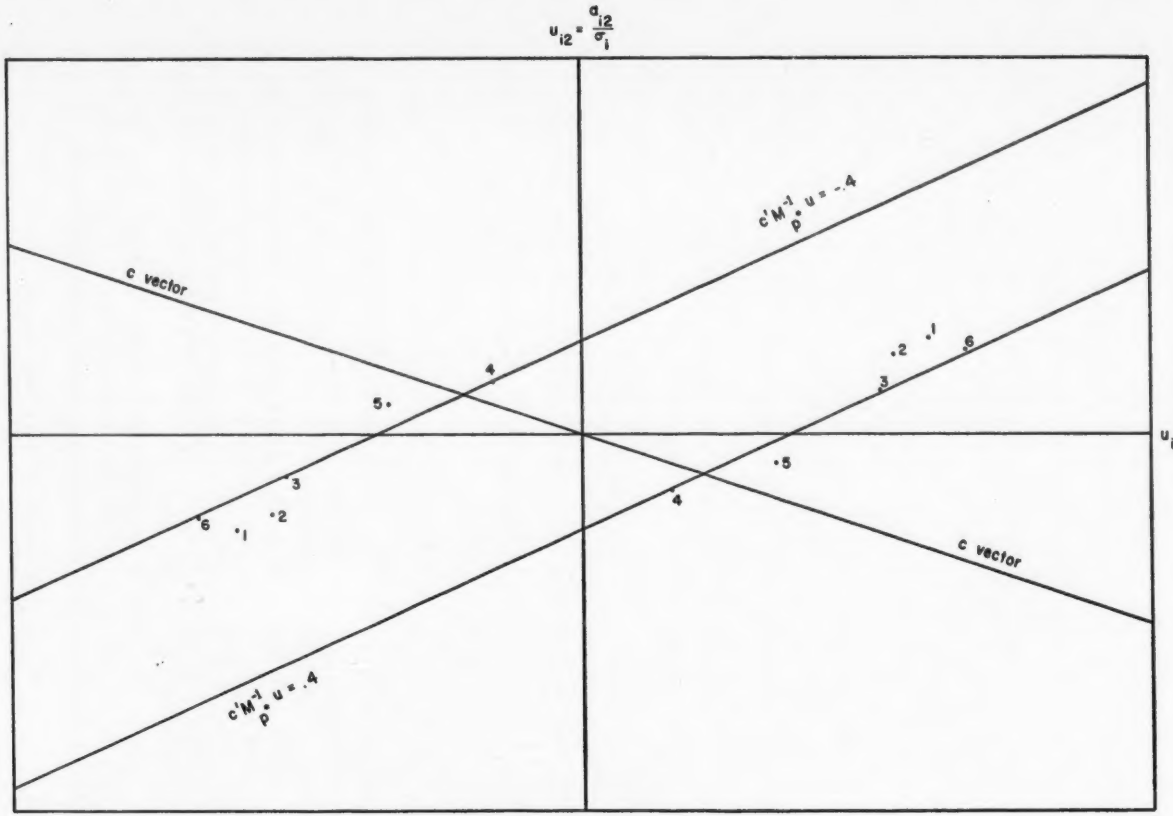

FIGURE 1
Illustration of Solution for Realistic Data: Section 4

Determine r such that

$$c'M_{p^*}^{-1}u_3 = c'M_{p^*}^{-1}u_4 .$$

That is,

$$\begin{aligned}(46)\qquad &[.641 \quad -.205]\begin{bmatrix} .118 - .012r & .345 - .511r \\ .345 - .511r & 1.079 + 2.202r \end{bmatrix}\begin{bmatrix} 1.588 \\ .231 \end{bmatrix} \\ &= [.641 \quad -.205]\begin{bmatrix} .118 - .012r & .345 - .511r \\ .345 - .511r & 1.079 + 2.202r \end{bmatrix}\begin{bmatrix} .566 \\ -.254 \end{bmatrix}.\end{aligned}$$

From this equation $r = .019$.

It follows that

$$(47)\qquad p^* = (0 \quad 0 \quad .019 \quad .981 \quad 1 \quad 0);$$

$$(48)\qquad M_{p^*} = \begin{bmatrix} 1.1208 & -.3353 \\ -.3353 & .1182 \end{bmatrix};$$

$$(49)\qquad c'M_{p^*}^{-1} = (.3506 \quad -.7399);$$

$$(50)\qquad \begin{aligned} &c'M_{p^*}^{-1}u_1 = .273, \quad c'M_{p^*}^{-1}u_2 = .286, \quad c'M_{p^*}^{-1}u_3 = .386, \\ &c'M_{p^*}^{-1}u_4 = .386, \quad c'M_{p^*}^{-1}u_5 = .476, \quad c'M_{p^*}^{-1}u_6 = .341. \end{aligned}$$

Thus, the conditions of the theorem are satisfied, and p^* is the minimizing vector. The variance of the associated estimate is .376. When this is contrasted with .381, the variance when item 4 and item 5 are used totally, a slight improvement by the use of fractional allocation is seen.

It should be remembered, however, that the modification of the problem by admitting fractional allocation is an ad hoc device, applying primarily when n is considerably larger than k; its justification is to be found in the paragraph following Theorem 1. Thus, in the present example, the *practical* problem is to find the best *integral* solution, i.e., the best pair of items. The fractional solution (47) suggests that items 4 and 5 comprise the best choice, and this is actually the case, as has been checked by comparing the variances corresponding to all possible choices of two items.

An Artificial Example when $k = 2$

For the selection of two out of six items, in the previous example the variance of z actually was computed for all 15 pairs in the two-factor structure; it was found that one could improve on the best pair by fractional allocation. It was shown how Theorem 1, combined with graphical considerations, could be used to select the best pair of items without going through all possible pairs. This becomes especially important when N and n both increase and thus rule out the examination of all possibilities. It also becomes important when k, the number of common factors, increases.

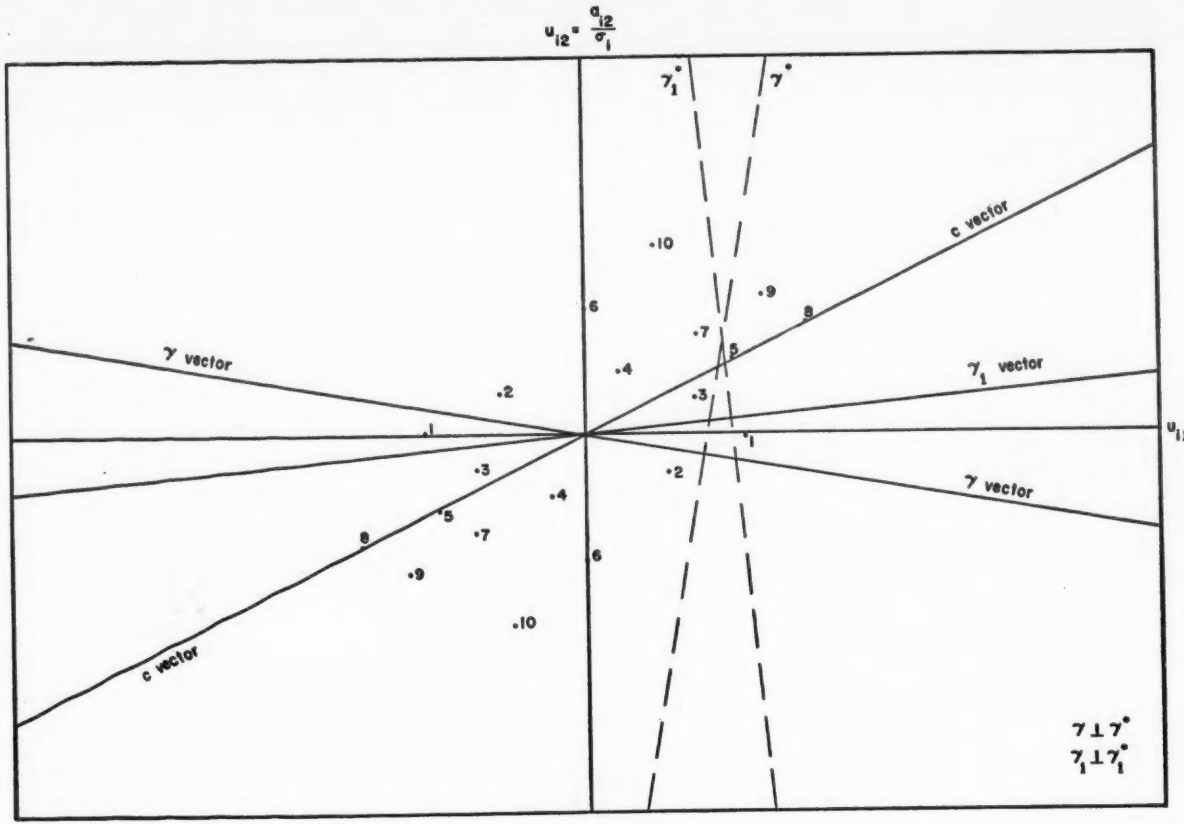

FIGURE 2
Illustration of Solution for Artificial Data: Section 5

The application of the theorem in a more complex situation will be demonstrated by the following artificial example. The problem is to obtain from the 10 items, first the optimal sets of 4 items, then of 5 items, and finally of 6 items. A common factor space of two dimensions is assumed. Thus,

$$x_i = a_i'y + \epsilon_i \qquad (i = 1, \cdots, 10), \tag{51}$$
$$z = c'y + \eta,$$

where $c' = (2 \quad 1)$. Assume

$$\begin{aligned} u_1' &= (\ 4 \quad 0), & u_6' &= (0 \quad 3), \\ u_2' &= (-2 \quad 1), & u_7' &= (3 \quad 3), \\ u_3' &= (\ 3 \quad 1), & u_8' &= (6 \quad 3), \\ u_4' &= (\ 1 \quad 2), & u_9' &= (5 \quad 4), \\ u_5' &= (\ 4 \quad 2), & u_{10}' &= (2 \quad 5); \end{aligned} \tag{52}$$

$$\begin{aligned} u_1u_1' &= \begin{bmatrix} 16 & 0 \\ 0 & 0 \end{bmatrix}, & u_6u_6' &= \begin{bmatrix} 0 & 0 \\ 0 & 9 \end{bmatrix}, \\ u_2u_2' &= \begin{bmatrix} 4 & -2 \\ -2 & 1 \end{bmatrix}, & u_7u_7' &= \begin{bmatrix} 9 & 9 \\ 9 & 9 \end{bmatrix}, \\ u_3u_3' &= \begin{bmatrix} 9 & 3 \\ 3 & 1 \end{bmatrix}, & u_8u_8' &= \begin{bmatrix} 36 & 18 \\ 18 & 9 \end{bmatrix}, \\ u_4u_4' &= \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}, & u_9u_9' &= \begin{bmatrix} 25 & 20 \\ 20 & 16 \end{bmatrix}, \\ u_5u_5' &= \begin{bmatrix} 16 & 8 \\ 8 & 4 \end{bmatrix}, & u_{10}u_{10}' &= \begin{bmatrix} 4 & 10 \\ 10 & 25 \end{bmatrix}. \end{aligned} \tag{53}$$

Applying the approximate method for complex situations, presented earlier,

$$\Lambda = \sum_{i=1}^{10} u_iu_i' = \begin{bmatrix} 120 & 68 \\ 67 & 88 \end{bmatrix}. \tag{54}$$

Then

$$\Lambda^{-1} \sim \begin{bmatrix} 78 & -68 \\ -68 & 120 \end{bmatrix} \tag{55}$$

and

$$\gamma' = c'\Lambda^{-1} \sim (88 \quad -16). \tag{56}$$

In (55) and (56) we need not be concerned with proportionality factors.

The item points are plotted graphically in Figure 2. Plotting the direc-

tion of γ in Figure 2, and imagining a straight line perpendicular to it moving from right to left, then the order in which the item points are swept over by this line indicates their order of preference in the first approximation.

If $n = 4$, i.e., 4 items are to be selected, one is led to the selection I_1 : 1, 5, 8, 9. Forming the corresponding product sums,

$$M_1^{-1} \sim \begin{bmatrix} 29 & -46 \\ -46 & 93 \end{bmatrix}, \qquad \gamma_1' \sim (12 \quad 1). \tag{57}$$

In the direction of γ_1 , the four outermost item points are still numbers 1, 5, 8, 9. The theorem shows, then, that choice I is optimal.

Next, take $n = 5$. Using the approximate selection determined by the γ-direction, one tries the set II_1 : 1, 3, 5, 8, 9. From the corresponding product sums, $\gamma_1' \sim (11 \quad 4)$. In this direction, however, the five outermost item points are II_2 : 1, 5, 7, 8, 9. On the other hand, the latter set yields the direction $\gamma_2' \sim (21 \quad -8)$, and the five outermost points 1, 3, 5, 8, 9, i.e., the set first attempted. This mutuality makes it plausible that a selection of form II_3 : 1, 3*, 5, 7*, 8, 9 (* indicating fractional selection) might give an optimal solution of the problem. Introducing unknown weights p_3 , p_7 with sum one,

$$M_3 \sim \begin{bmatrix} 93 + 9p_3 + 9p_7 & 46 + 3p_3 + 9p_7 \\ 46 + 3p_3 + 9p_7 & 29 + p_3 + 9p_7 \end{bmatrix}. \tag{58}$$

In order for the points 3, 7 to lie on the boundary line of the selection region, the vector $\gamma = M^{-1}c$ must have direction (1, 0), which gives the condition

$$-2(46 + 3p_3 + 9p_7) + (93 + 9p_3 + 9p_7) = 0; \tag{59}$$

hence (noting that $p_3 + p_7 = 1$), $p_3 = 2/3$, $p_7 = 1/3$. Thus the optimum allocation vector is $p^* = (1, 0, 2/3, 0, 1, 0, 1/3, 1, 1, 0)$. In practice, one must make a choice between selections II_1 and II_2 . The corresponding variances $V = c'M^{-1}c$ are (note that the factors $1/|M|$ cannot be omitted at this point)

$$V_1 = \frac{30(2^2) - 2(49)(2)(1) + 102(1^2)}{102(30) - 49(49)} = \frac{26}{659} = 0.03945,$$

$$V_2 = \frac{38(2^2) - 2(55)(2)(1) + 102(1^2)}{102(38) - 55(55)} = \frac{34}{851} = 0.03995.$$

Thus, there will be a slight preference for the former choice.

Finally, if one wishes to include $n = 6$ items, it seems reasonable, from the figure, to try the set III: 1, 3, 5, 7, 8, 9. It leads to $\gamma_1' \sim (20 \quad -5)$. In this case, the moving boundary line will hit points 2 and 7 at the same time. Nevertheless, the set III still fulfills the condition of the theorem, and hence is optimal. Note that in this example, of the two item points on the boundary line, item 7 has weight 1, and item 2 has weight 0.

Early in the paper it was stated that the optimal $(n + 1)$ set of items need not completely contain the optimal set of n items. However in many practical situations this will occur and the present exercise demonstrates this event.

REFERENCES

[1] Elfving, G. Selection of item variables for prediction. Report No. 56-91. Air Univ., School of Aviation Medicine, USAF, Randolph AFB, Texas, August, 1956.

[2] Elfving, G. Further contributions to the theory of item selection. Report No. 57-97. Air Univ., School of Aviation Medicine, USAF, Randolph AFB, Texas, June, 1957.

[3] Elfving, G. Optimum allocation in linear regression theory. *Ann. math. Statist.*, 1952, **23**, 225-262.

[4] Elfving, G. Geometric allocation theory. *Skandinavisk Aktuarietidskrift.*, 1954, **37**, 170-190.

[5] Elfving, G. A selection problem in experimental design, *Societas Scientarium Fennica, Commentationes Physico-Mathematicae*, 1957, **20**, Ch. 2, pp. 1-10.

[6] Elfving, G. Selection of nonrepeatable observations for estimation. In J. Neyman (Ed.), *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley: Univ. California Press, 1955. Pp. 69-75.

Manuscript received 8/18/58

PSYCHOMETRIKA—VOL. 24, NO. 3
SEPTEMBER, 1959

A METRIC AND AN ORDERING ON SETS

FRANK RESTLE

MICHIGAN STATE UNIVERSITY

Beginning with sets of arbitrary elements, concepts of distance and betweenness of sets are defined. Since betweenness as defined is not transitive, an investigation is made of the conditions which ensure desirable regularity. It is found that a straight line or linear array of sets is a generalization of nested sets (Guttman scales). Close relationships among the notions of distance, betweenness, and linear arrays are demonstrated. Parallel and perpendicular arrays, dimensions, and multidimensional spaces are characterized.

Basic to many psychological discussions is the concept of similarity, which is used to arrange objects or events. Two things which are quite similar are psychologically close together, and two things which are quite dissimilar are psychologically distant. This interpretation of dissimilarity as distance is the basis of numerous attempts to give a simplifying, quantitative structure to psychological data in psychophysics, learning, and other fields.

In contrast to these metric, quantitative developments are more qualitative discussions of similarity and the arrangement of objects, using the idea of common elements to account for similarity. In learning, such views are common in current mathematical theories [3]. Some recent efforts have been made to generate a metric analysis from the set-theoretic considerations of learning theory [2, 3] and psychophysics [4]. The present paper shows how a distance and ordering can be defined on sets, and generates the beginnings of the resulting geometry.

The arbitrary elements to be considered have no individual locations or arrangement except with regard to the sets of which they are members. Thus, the foundation is entirely qualitative. The purpose is to determine what internal evidence would justify saying that a sequence of sets is ordered, and to construe a concept of *distance* between sets. The conceptual apparatus used is elementary set theory.

The following mathematical concepts and notations are used.

$\mathfrak{U}$ the universe.

$\emptyset$ the empty set.

$\cup$ union or set sum; $A \cup B$ is the set containing everything in either A or B or both.

$\cap$ intersection or set product; $A \cap B$ is the set containing just those elements common to both A and B.

- set-complement; $\bar{A}$ is the set containing all elements of $\mathfrak{U}$ which are not in A.

$\subset$ the relation of inclusion; $A \subset B$ if and only if all members of A are also members of B.

m a measure function; if S^* is a set of sets $\{S_1, S_2, \cdots\}$ and m is a measure function, then the following axioms hold.

$M1$: for all S_i, $m(S_i) \geq 0$.

$M2$: $m(\phi) = 0$.

$M3$: if $S_i \cap S_j = \phi$, then $m(S_i \cup S_j) = m(S_i) + m(S_j)$.

$\{a, b\}$ is the set whose only members are a and b. In general, entities enclosed in curved brackets are the members of a set.

$\langle a, b \rangle$ the ordered set whose first member is a and second member is b. In general, entities enclosed in angle brackets are members, in the indicated order, of an ordered set. Two ordered sets are identical only if they have the same elements in the same order. (Unordered sets are identical if they have the same members—order is immaterial.)

Consider a universe $\mathfrak{U}$ of arbitrary elements, which may be interpreted as the universe of stimulus elements, cues, etc., in learning or perceptual models, or which might also be the set of responses to test items, or the population of subjects, etc., in other applications.

Certain subsets of $\mathfrak{U}$ can be distinguished by the kind of observation used in the particular experiment. Therefore consider a set $\mathfrak{S} = \{S_1, S_2, \cdots, S_n\}$ of subsets of $\mathfrak{U}$. A necessary addition is a notion of magnitude of a set, which may be the number of elements in the set, or (if the elements receive different weights) the sum of weights of elements, or which may be a more abstract idea of magnitude. The function m is a measure function on the sets in $\mathfrak{S}$, and also on any sets which can be formed from members of $\mathfrak{S}$ by the operations of union, intersection, and complementation performed a countable number of times. Let this larger set be $\mathfrak{S}^*$. This ensures that m will be defined on all the sets in which we have interest.

Metric of Dissimilarity

Let the dissimilarity or *distance* between two sets S_i and S_j be written D_{ij}. For D to be a metric distance, it must satisfy the following metric axioms.

AXIOM 1. $D_{ij} = 0$ if and only if $i = j$.

AXIOM 2. $D_{ij} \geq 0$.

AXIOM 3. $D_{ij} = D_{ji}$.

AXIOM 4. $D_{ij} + D_{jk} \geq D_{ik}$.

Within the present system, the distance D_{ij} must be defined in terms of the sets S_i and S_j, combinations produced by intersection, union, and comple-

mentation, and the measure function m. The nature of the elements in $\mathfrak{U}$ does not in general give a concept of distance between elements upon which to build the concept of distance between sets.

Other things equal, the degree of dissimilarity between two sets depends on the measure of noncommon elements, i.e., the symmetric set difference. The first proposal of this study is to let the measure of the symmetric set difference be the distance metric.

DEFINITION 1. $D_{ij} = m\,[(S_i \cup S_j) \cap \overline{(S_i \cap S_j)}]$.

The expression on the right is, of course, the symmetric set difference.

THEOREM 1. *D_{ij} is a metric.*

PROOF. $S_i \cup S_i = S_i \cap S_i = S_i$, whence $D_{ii} = m(S_i \cap \bar{S}_i) = m(\phi) = 0$. If m vanishes only for the empty set, D satisfies Axiom 1 of a metric function. Axiom 2 is satisfied because measure functions have nonnegative values, and Axiom 3 is satisfied because of the symmetry of the expression in Definition 1. Axiom 4, the triangle inequality, is shown to hold by partitioning $\mathfrak{U}$ into the following eight cells (see Figure 1):

$$S_i \cap S_j \cap S_k = A_1\,; \qquad S_i \cap S_j \cap \bar{S}_k = A_2\,;$$
$$S_i \cap \bar{S}_j \cap S_k = A_3\,; \qquad \bar{S}_i \cap S_j \cap S_k = A_4\,;$$
$$S_i \cap \bar{S}_j \cap \bar{S}_k = A_5\,; \qquad \bar{S}_i \cap S_j \cap \bar{S}_k = A_6\,;$$
$$\bar{S}_i \cap \bar{S}_j \cap S_k = A_7\,; \qquad \bar{S}_i \cap \bar{S}_j \cap \bar{S}_k = A_8\,.$$

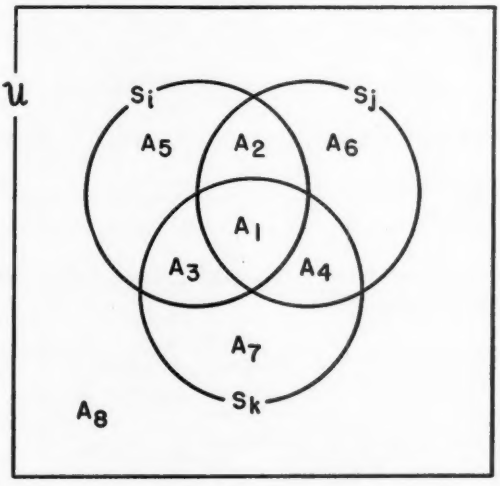

FIGURE 1

Three sets, S_i, S_j, and S_k, in a universe $\mathfrak{U}$, with the cells of the resulting partition

For compactness in what follows, let $m(A_i) = a_i$. Now,

$$D_{ij} = a_3 + a_4 + a_5 + a_6 ,$$

$$D_{jk} = a_2 + a_3 + a_6 + a_7 ,$$

$$D_{ik} = a_2 + a_4 + a_5 + a_7 ;$$

whence

$$D_{ij} + D_{jk} = D_{ik} + 2a_3 + 2a_6 ,$$

and

$$D_{ij} + D_{jk} \geq D_{ik} .$$ Q.E.D.

An Ordering on Sets

Since the elements of $\mathfrak{U}$ are arbitrary and not already located in a space, or otherwise ordered, an ordering of sets must be defined in terms of the membership of the sets. It appears better to be unrealistically restrictive than vague in the present analysis.

Consider first what it means to say that S_j is *between* S_i and S_k . Two conditions seem sufficient to justify such a description; first that S_i and S_k have no common members which are not also in S_j , and second, that S_j have no unique members which are in neither S_i nor S_k . Figure 2a shows an S_j which is between S_i and S_k in this sense. Figure 2b shows an S_j' which fails to meet the first condition. Figure 2c shows an S_j'' which fails to meet the second condition.

This concept can be written as the following definition.

DEFINITION 2. S_j is between S_i and S_k (written b_{ijk}) if and only if

(i) $S_i \cap \bar{S}_j \cap S_k = \emptyset$ (A_3 in Figure 1 is empty),
(ii) $\bar{S}_i \cap S_j \cap \bar{S}_k = \emptyset$ (A_6 in Figure 1 is empty).

A major justification of definitions 1 (the metric) and 2 (betweenness) is found in the following theorem.

THEOREM 2. *If b_{ijk} then $D_{ij} + D_{jk} = D_{ik}$; if m vanishes only for the empty set then the converse is true.*

PROOF. From the proof of Theorem 1,

$$\begin{aligned} D_{ij} + D_{jk} &= D_{ik} + 2a_3 + 2a_6 \\ &= D_{ik} + 2m(S_i \cap \bar{S}_j \cap S_k) + 2m(\bar{S}_i \cap S_j \cap \bar{S}_k). \end{aligned}$$

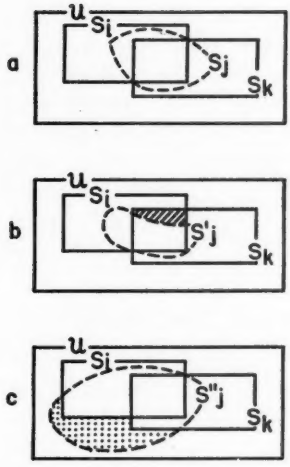

FIGURE 2

An example of ordered sets (*a*) and two examples of unordered sets (*b*) and (*c*)

By hypothesis, b_{ijk}, the sets $S_i \cap \bar{S}_j \cap S_k$ and $\bar{S}_i \cap S_j \cap \bar{S}_k$ are empty, whence $D_{ij} + D_{jk} = D_{ik}$. The converse is obvious if $a_3 = a_6 = 0$ implies that A_3 and A_6 are empty. Q.E.D.

Thus, b_{ijk} means that, in an abstract sense, S_j lies on a straight line between S_i and S_k, in that a path from S_i to S_k through S_j is just as short as any path. A preliminary investigation has shown, however, that one does not easily build a rigid metric space from the relation b. The main trouble is that b is not transitive in what would be a most useful way. The problems are shown in the two counter examples proving Theorem 3.

THEOREM 3. *It is not the case that b_{ijk} and b_{jkm} implies b_{ijm}.*

PROOF. Two counter examples are given because of their intrinsic interest.

First counter example: let $S_i = \{1, 2, 3\}$, $S_j = \{2, 3, 4\}$, $S_k = \{3, 4, 5\}$, and $S_m = \{4, 5, 1\}$. Then b_{ijk} and b_{jkm}, but not b_{ijm} because $S_i \cap \bar{S}_j \cap S_m = \{1\}$, whereas it must be empty for b_{ijm} to hold.

Second counter example: let $S_i = \{1, 2\}$, $S_j = \{2, 3\}$, $S_k = \{3, 4\}$ and $S_m = \{4, 5\}$. Then b_{ijk} and b_{jkm}, but not b_{ijm} because $\bar{S}_i \cap S_j \cap \bar{S}_m = \{3\}$, whereas it must be empty for b_{ijm} to hold. Q.E.D.

These two examples correspond to what may be considered psychological reality. The first counter example shows a case of *looping* in which a series of neighboring sets, each between its two neighbors, may lead around and back to the starting set. This is the apparent structure of the color circle, the

array of hues from red to yellow to green to blue to purple and back to red. In the second counter example, one detours from S_i to S_m through some sets which may be quite unlike either. For example, a series of drawings might be made, varying from starkly simple to profusely ornate. Drawings in the middle of the series could have psychological properties, such as attractiveness, shared by neither extreme.

Since a long straight line, or linear array, of sets cannot be built up from shorter segments in a reliable way, for use as a definition of a linear array let us seek a reasonable general description of the conditions under which loops and detours can be avoided.

The Special Case of Nested Sets—The Guttman Scale

Consider a finite sequence of sets S_1 , S_2 , $\cdots$, S_n such that $S_i \subset S_{i+1}$ for $i = 1, 2, \cdots, n$. These sets are nested within one another with S_n being the largest. Let such a sequence be called N.

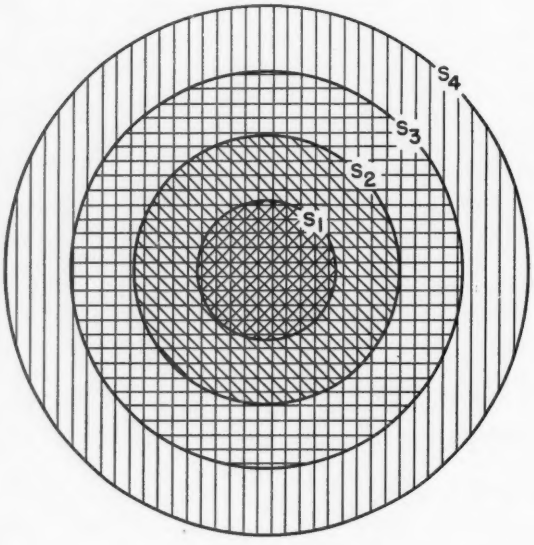

FIGURE 3
A nested sequence of sets as referred to in Theorem 4

THEOREM 4. *If S_i , S_j , and S_k are in N and $i \leq j \leq k$, then b_{ijk} .*

PROOF. $S_i \cap \bar{S}_j \cap S_k$ is a subset of $S_i \cap \bar{S}_j$ which is empty, and $\bar{S}_i \cap S_j \cap \bar{S}_k$ is a subset of $S_j \cap \bar{S}_k$ which is empty, establishing b_{ijk} . Q.E.D.

COROLLARY. *If S_i , S_j , and S_k are in N and $i \leq j \leq k$, then $D_{ij} + D_{jk} = D_{ik}$.*

At least up to this point, the material given is familiar and found in standard mathematical texts (e.g., [7], pp. 16, 24, 107, Exercise g).

The condition that an array of sets be nested is sufficient to ensure transitivity of betweenness and additivity of distances, but it seems quite restrictive in that it certainly does not correspond to qualitative scales. However, it is possible to make a qualitative or substitutive scale from two nested arrays and a constant set.

A Linear Array of Sets

The following development, like that above, will be restricted to finite arrays of sets. No great difficulties are foreseen in generalizing, but the ability to deal with finite numbers of sets is an advantage in empirical applications of the system.

DEFINITION 3. Let $A^* = \langle A_1 , A_2 , \cdots , A_n \rangle$ and $B^* = \langle B_1 , B_2 , \cdots , B_n \rangle$ be two nested n-tuples of sets, with the restriction that $A_n \cap B_n = \phi$ (whence all the A_i are disjunct from all the B_j). Let C be any set in S^* such that $A_n \cap C = B_n \cap C = \phi$. Then the n-tuple of sets

$$L^* = \langle L_1 , L_2 , \cdots , L_n \rangle,$$

where

$$L_i = A_i \cup B_{n-i+1} \cup C,$$

is a *linear array of sets.*

THEOREM 5. *If L_i , L_j , and L_k are in a linear array L^* , and $i \leq j \leq k$, then b_{ijk} .*

PROOF. Partition L_i into A_i , B_{n-i+1} , and C.

By the hypothesis that A^* is a nested sequence, and Theorem 4,

$$A_i \cap \bar{A}_j \cap A_k = \phi.$$

Similarly,

$$B_{n-i+1} \cap \bar{B}_{n-j+1} \cap B_{n-k+1} = \phi,$$

and certainly

$$C \cap \bar{C} \cap C = \phi.$$

Now,

$$L_i \cap \bar{L}_j \cap L_k = (A_i \cup B_{n-i+1} \cup C) \cap \overline{(A_j \cup B_{n-j+1} \cup C)} \cap (A_k \cup B_{n-k+1} \cup C)$$

which, because the A's, B's and C are all disjoint, is equal to

$$(A_i \cap \bar{A}_j \cap A_k) \cup (B_{n-i+1} \cap \bar{B}_{n-j+1} \cap B_{n-k+1}) \cup (C \cap \bar{C} \cap C) = \emptyset.$$

Similar argument shows that

$$\bar{L}_i \cap L_j \cap \bar{L}_k = \emptyset. \qquad \text{Q.E.D.}$$

COROLLARY. *If L^* is a linear array of sets and $i \leq j \leq k$, then*

$$D_{ij} + D_{jk} = D_{ik} .$$

PROOF. Theorems 5 and 2.

Note that nested sets are a special case of a linear array in which, for example, $B_i = \emptyset$ for $i = 1, 2, \cdots, n$. Therefore, Theorem 4 is a special case of Theorem 5.

The restrictions of a *linear array* (Definition 3) have been shown to be sufficient to ensure transitivity of betweenness. These same restrictions will now be shown to be a necessary condition for such transitivity.

THEOREM 6. *If $R^* = \langle R_1 , R_2 , \cdots , R_n \rangle$ is a sequence of sets such that for all $i, j, k, = 1, 2, \cdots, n$, and $i \leq j \leq k$, b_{ijk} , then R^* is a linear array of sets.*

PROOF. It must be shown that there exist nested sequences $A^* = \langle A_1 , \cdots , A_n \rangle$ and $B^* = \langle B_1 , \cdots , B_n \rangle$ and a set C such that $R_i = A_i \cup B_{n-i+1} \cup C$, and $A_n \cap B_n = \emptyset$, $A_n \cap C = \emptyset$ and $B_n \cap C = \emptyset$.

(i) Define, for an example,

$$C = R_1 \cap R_2 \cap \cdots \cap R_n \qquad (= R_1 \cap R_n),$$

$$A_i = R_i \cap \bar{R}_1 ,$$

$$B_j = R_{n-j+1} \cap \bar{R}_n .$$

(ii) Show that $A_i \subset A_{i+1}$. This is equivalent to the statement that

$$(R_i \cap \bar{R}_1) \subset (R_{i+1} \cap \bar{R}_1),$$

or

$$(R_i \cap \bar{R}_1) \cap \overline{(R_{i+1} \cap \bar{R}_1)} = \emptyset,$$

or

$$(R_i \cap \bar{R}_1) \cap (\bar{R}_{i+1} \cup R_1) = \emptyset.$$

Multiplying,

$$(R_i \cap \bar{R}_1 \cap \bar{R}_{i+1}) \cup (R_i \cap \bar{R}_1 \cap R_1) = \emptyset,$$

which is true: the first term is empty by $b_{1,i,i+1}$, and the second term is obviously empty. Similar steps show that the B's are nested.

(iii) Show that $A_i \cup B_{n-i+1} \cup C = R_i$.

$$A_i \cup B_{n-i+1} \cup C = (R_i \cap \bar{R}_1) \cup (R_{n-(n-i+1)+1} \cap \bar{R}_n) \cup (R_1 \cap R_n)$$
$$= (R_i \cap \bar{R}_1) \cup (R_i \cap \bar{R}_n) \cup (R_1 \cap R_n).$$

Factoring the first two terms gives

$$A_i \cup B_{n-i+1} \cup C = [R_i \cap (\bar{R}_1 \cup \bar{R}_n)] \cup (R_1 \cap R_n)$$
$$= [R_i \cap (\overline{R_1 \cap R_n})] \cup (R_1 \cap R_n).$$

By the hypothesis b_{1in} , it is seen that $R_1 \cap R_n = R_i \cap (R_1 \cap R_n)$. Substituting this in the above equation,

$$A_i \cup B_{n-i+1} \cup C = [R_i \cap (\overline{R_1 \cap R_n})] \cup [R_i \cap (R_1 \cap R_n)]$$
$$= R_i .$$

The rest of the proof is trivial.

Theorem 6 shows, in a formal way, that any sequence of sets throughout which the relation of betweenness holds may be dissected into two nested sequences, running in opposite directions, and a constant remainder. This gives a sharp meaning to the concept of a substitutive scale, for as one moves from R_i to R_{i+1} , part of the elements of B_n are removed, and some new elements from A_n are substituted. There is no requirement that the measure added must equal the amount removed, though this may be an interesting special case.

From Theorem 6 it is seen that a linear array of sets is well represented by three generic sets, A^*, B^*, and C. A_n and B_n will be called the *poles*, the nested sequences A^* and B^* the *polar arrays*, and C the *core*, it being common to all elements in the original array.

Multidimensional Arrays

In current multidimensional scaling it is common to assume an arrangement in Euclidean space. Attneave [1] and Landahl [6] have suggested non-Euclidean rules for computing the distances between objects which vary on more than one dimension, suggesting that psychological intuition does not force us to the Euclidean assumption. The present discussion suggests a generalization of the above system to more than one dimension, giving rise to a non-Euclidean arrangement like those mentioned above.

DEFINITION 4. Two linear arrays of sets, L^* and L'^*, are said to be *parallel* if and only if their polar arrays are identical.

LEMMA 1. *If L^* and L'^* are parallel arrays, $D_{ij} = D_{i'j'}$.*

PROOF. This follows at once from the fact that distance depends only on the polar sets, not at all on the core (or common elements).

LEMMA 2. *If L^* and L'^* are parallel linear arrays,*

$$D_{L_iL_{i'}} = D_{ii'} = D_{ii} + D_{C,C'} .$$

PROOF. Partition the universe $\mathfrak{U}$ into the following cells:

$$Q_1 = L_i \cap L_i' \cap C \cap C'; \qquad Q_2 = L_i \cap \bar{L}_i' \cap \bar{C} \cap \bar{C}';$$
$$Q_3 = \bar{L}_i \cap L_i' \cap \bar{C} \cap C'; \qquad Q_4 = L_i \cap \bar{L}_i' \cap C \cap \bar{C}';$$
$$Q_5 = L_i \cap \bar{L}_i' \cap \bar{C} \cap \bar{C}'; \qquad Q_6 = \bar{L}_i \cap L_i' \cap \bar{C} \cap \bar{C}';$$
$$Q_7 = \bar{L}_i \cap \bar{L}_i' \cap \bar{C} \cap \bar{C}'.$$

The cells are clearly mutually exclusive. They are exhaustive because seven other possible cells are subsets of $C \cap \bar{L}_i$ (empty) or $C' \cap \bar{L}_i'$ (also empty), and the other two possible cells, $L_i \cap L_i' \cap \bar{C} \cap C'$ and $L_i \cap L_i' \cap C \cap \bar{C}'$ are identical with Q_2 , being comprised of the elements common to L_i and L_i' which are not in the cores. A diagram of the sets and the partition is shown in Figure 4. Here, $D_{ii'} = m(Q_5) + m(Q_6)$, $D_{C,C'} = m(Q_3) + m(Q_4)$,

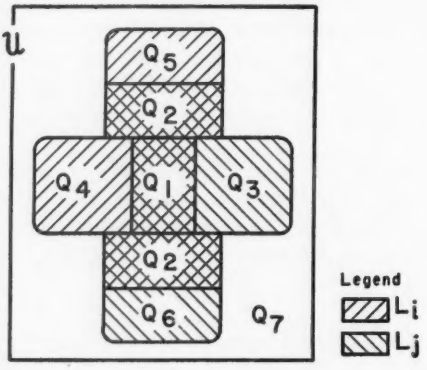

FIGURE 4
The sets L_i and L_i' and the partition referred to in Lemma 2

and as stated by the lemma, $D_{ii'} = m(Q_3) + m(Q_4) + m(Q_5) + m(Q_6) = D_{ii} + D_{CC'}$. Q.E.D.

In what follows, a set of parallel linear arrays will be referred to as a *dimension*, using the following notation. A dimension X_L^* is a set of linear arrays, $L^{*(1)}, L^{*(2)}, \cdots, L^{*(t)}$, which are parallel to one another.

DEFINITION 5. If $X_L^* = L^{*(1)}, \cdots, L^{*(t)}$ is a dimension and the cores of the arrays, $C^{(1)}, \cdots, C^{(t)}$ form a linear array, then the set of members of $L^{*(1)}, \cdots, L^{*(t)}$ is called a *linear two-space*.

THEOREM 7. *If S^* is a linear two-space made up of the linear arrays*

$$\begin{aligned} L^{*(1)} &= \langle L_1^{(1)}, \cdots, L_r^{(1)}, \cdots, L_n^{(1)} \rangle, \\ &\vdots \quad\quad \vdots \quad\quad\quad \vdots \quad\quad\quad \vdots \\ L^{*(\ell)} &= \langle L_1^{(\ell)}, \cdots, L_r^{(\ell)}, \cdots, L_n^{(\ell)} \rangle, \end{aligned}$$

then the sequence of sets composed of the rth member of each array, i.e., $L_r^{(1)}, \cdots, L_r^{(i)}, \cdots, L_r^{(\ell)}$ is also a linear array.

PROOF. By definition of a two-space the cores form a linear array, and by adding to each core the constant set $L_r^{(i)} \cap \overline{C^{(i)}}$, the result is still a linear array. Q.E.D.

Theorem 7 shows that there is no asymmetry between the two dimensions of a two-space. If X^* and Y^* are the two dimensions, the cores of X^* give the polar arrays of Y^*, and the cores of Y^* give the polar arrays of X^*.

Further dimensions may be constructed by considering the joint core of a two-space, i.e., the intersection of all sets in the space. If one were to construct a sequence of parallel two-spaces, identical except for their joint cores, and if the resulting sequence of joint cores were a linear array, then one would have a three-space. This process can be repeated indefinitely, so that an n-space is defined for any n.

A Non-Pythagorean Theorem, and Description of the Space

An advantage of Euclidean space, which has contributed to its use in psychological measurement, is the Pythagorean theorem, that the distance between two points which are apart, on various dimensions, by amounts $x_1, x_2, \cdots, x_n$, is given by $d = \sqrt{\sum_n x_i^2}$. If two sets in an n-space of the present type are apart, on the various dimensions, by amounts $x_1, x_2, \cdots, x_n$, then the distance between them is $d = \sum_n x_i$. This is shown informally by applying Lemmas 1 and 2. The only differential elements must, by definition of an n-space, be elements in the linear arrays in the n dimensions. Since the poles of the various dimensions are discrete, the measures are all added. That the result is unique follows from Lemma 2 and the symmetry of dimensions.

One gets an idea of the space by considering a small finite two-space, rectangular in its Euclidean mapping, like that in Figure 5. The sets are represented by points, and the *paths* from point a to point b are indicated.

The number of paths from a to b in any two-space is easily computed. If a and b are separated by r sets in one dimension and by s sets in the other, the number of paths is $\binom{r+s}{r}$. Of course, the steps need not be all of the same distance, but the total lengths of all paths will be the same.

Along with the concepts of parallel arrays, and perpendicular arrays (in which the poles have no overlap), one can also consider oblique arrays

(in which there is some overlap between the poles). In Figure 5, with the particular space of sets available, one proceeds from a to b by changes in one dimension at a time. It would be possible, if sets can be produced experimentally (as where they are physical objects in a perception experiment) to introduce new sets whereby one goes directly from a to b by successive changes in *both* dimensions simultaneously. The total distance will be the same, of course, because one either introduces more intermediate sets or has

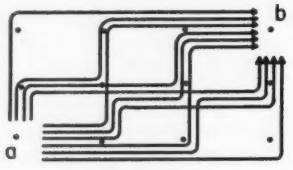

FIGURE 5
A two-space with 3 x 4 sets, each mapped on a point
(The ten ways of going from a to b are indicated.)

sets which, because they differ in two dimensions, are farther apart than the ones shown in Figure 5. However, this possibility shows that the present geometry does not permit formal identification of *pure variables*. If one had two pure psychological variables, each of which gives rise to linear arrays, then a new linear array could be made by uniting the first two, element by element (provided, of course, that they are of the same number of elements). This new compound array would have all the formal properties of a linear array and would be indistinguishable from a pure array in this system. This same fact applies to Guttman scales as a special case and is presumably rather general, suggesting that a pure variable cannot be isolated by internal evidence within this system.

Discussion

The similarity between the axioms of a measure function and the metric axioms has been used to devise a way of measuring the distance between sets. Hays [5] has used the same distance concept, calling it "implicational difference," and defining it by

$$\mathrm{Di}_{AB} = m(A) + m(B) - 2m(A \cap B),$$

where A and B are sets. In this paper distance is defined as the measure of the symmetric set difference, which is identical with Hays' formulation.

Certain related concepts of the distance between sets may be mentioned for contrast. Galanter [4] bases his distance measure on the ratio of the measure of the set difference to the measure of the union of the two sets. If this distance is called G_{AB}, it would be

$$G_{AB} = \frac{m(A) + m(B) - 2m(A \cap B)}{m(A) + m(B) - m(A \cap B)}.$$

This distance is restricted between zero and one and does not have the desirable property of additivity when sets are ordered. Bush and Mosteller [3] define a similarity index η as the ratio of the measure of the intersection to the measure of one of the sets. That is,

$$\eta(A \text{ to } B) = m(A \cap B)/m(A).$$

This index cannot be the basis of a distance since it is not generally symmetric. If one used the mean or sum of the two directional η's, the result would be similar to Galanter's proposal.

With respect to the ordering of sets, both Galanter [4] and Burke [2] have presented ideas like the present one, and the writer also previously proposed a similar ordering [8]. Burke distinguished between substitutive and nested orderings in essentially the way done here, but did not give a general treatment of substitutive orderings and ways of escaping the intransitivity of the local relationship. Hays [5] does not consider an ordering of the present type, and instead embeds the numerical distances in a Euclidean space. In the writer's opinion, this loses part of the advantage of the set-theoretic formulation.

There are occasions in psychological theory in which a dimensional or metric approach is entirely natural and valuable, as in dealing with loudness, pitch, or hue, in the discussion of generalization gradients in learning, in treating polarized attitudes, and in certain applications of statistics. On the other hand, the categorical approach has a firmer logical foundation and the advantage that raw data in psychological experiments are usually categorical. The present paper is an attempt to bridge the gap between the two approaches by showing a way to develop dimension from purely set-theoretic concepts. It is hoped that this may help in unifying the mathematical approaches to psychological problems.

REFERENCES

[1] Attneave, F. Dimensions of similarity. *Amer. J. Psychol.*, 1950, **63**, 516-556.

[2] Burke, C. J. Some problems in discrimination models. *Amer. Psychologist*, 1956, **11**, 483.

[3] Bush, R. R. and Mosteller, F. A model for stimulus generalization and discrimination. *Psychol. Rev.*, 1951, **58**, 413-423.

[4] Galanter, E. H. An axiomatic and experimental study of sensory order and measure. *Psychol. Rev.*, 1956, **63**, 16-28.

[5] Hays, W. L. An approach to the study of trait implication and trait similarity. In R. Tagiuri and L. Petrullo (Eds.), *Person perception and interpersonal behavior*. Stanford: Stanford Univ. Press, 1958.

[6] Landahl, H. D. Neural mechanisms for the concepts of difference and similarity. *Bull. math. Biophysics*, 1945, **7**, 83-88.

[7] Munroe, M. E. *Introduction to measure and integration.* Cambridge, Mass.: Addison Wesley, 1953.
[8] Restle, F. A theory of discrimination. Unpublished doctoral dissertation, Stanford Univ., 1953.

Manuscript received 10/2/58

Revised manuscript received 3/2/59

PSYCHOMETRIKA—VOL. 24, NO. 3
SEPTEMBER, 1959

MAXIMUM LIKELIHOOD ESTIMATES OF ITEM PARAMETERS USING THE LOGISTIC FUNCTION

A. E. MAXWELL

INSTITUTE OF PSYCHIATRY, LONDON UNIVERSITY

The logistic function is proposed as an alternative to the integrated normal function when estimating parameters of test items. The logistic curve is described; an iterative method for finding maximum likelihood estimates of its parameters is given, and an example of its use is presented.

Finney [9] pointed out that the problem of estimating the *limina* and *precisions* of test items, discussed earlier by Ferguson [8], was closely analogous to problems in toxicological experiments. In particular he showed that the methods of probit analysis could be applied with advantage to Ferguson's data to obtain efficient estimates of the parameters in question. More recently Berkson [2, 3] and Anscombe [1] have shown that in toxicological work a useful alternative to the probit model, in which an integrated normal response law is assumed, is one in which the logistic function is employed. They point out that while the shapes of these two types of response curve are very similar, simple sufficient statistics are available for the parameters of the logistic curve. Attracted by this property Birnbaum [5], while "considering the application of the Neyman-Pearson and Wald theories of inference and statistical decision making to problems of efficient design and use of tests of a single ability," demonstrated some advantages of adopting the logistic function, rather than the usual normal ogive, for ascertaining the probability that an examinee of given ability would have a specific response pattern to a k-item test. While Birnbaum's approach (also illustrated in [6] and [7]) should be viewed as a portent of the shape of future test theory, it is useful meanwhile to look at the logistic model as an alternative to the probit model for the straightforward estimation of the limina and precision of test items.

The Problem as Conventionally Stated

Conventionally (for references see [11]) it has been assumed that the probability P_{ik} that an examinee k, of "true" ability x_k on a test sampling an ability x, will answer the item j of the test correctly is given by the expression

$$P_{ik} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(x_k - a_j)/\sigma_j} e^{-u^2/2}\, du, \tag{1}$$

where α_i and σ_i are constants for that item. When P_{ik} is 0.5, x_k is equal to α_i so that, following psychophysical nomenclature, α_i is the *limen* of the item, that point on the x-scale below which 50 percent of the population fail the item. The constant σ_i determines how well the item discriminates between individuals of high and low ability. The problem is one of finding efficient estimates, $\hat{\alpha}_i$ and $\hat{\sigma}_i$, of the population parameters α_i and σ_i . For completeness special reference should be made here to Lord [11], where he points out that (1) can be obtained, if desired, from certain assumptions, prominent amongst which are

(*i*) that there is a continuous variable x underlying the item, and
(*ii*) that x is normally distributed.

Lord then shows that if the metric x is the common factor of a set of items, measuring a single ability, x_k is invariant no matter what test of the ability is administered, and that this invariance holds whether or not x is normally distributed in the group tested. It should be noted that the common factor referred to here is not determined by a rank-one matrix of interitem tetrachorics (except in the special case where the common factor is normally distributed). It is found by an analog of latent structure analysis.

The probit of the proportion P_{ik} is defined [9] as the value of Y satisfying the equation

$$P_{ik} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(Y-5)} e^{-u^2/2}\, du, \tag{2}$$

where Y is the unit normal deviate corresponding to P_{ik} increased by 5. The probit of P is then linearly related to x by the equation

$$Y = 5 + (x - \alpha_i)/\sigma_i\ , \tag{3}$$

and by the familiar methods of probit analysis the constants α_i and σ_i can be estimated.

The Logistic Model

Suppose that groups of individuals are tested at each of k different values of the variable x_i ($i = 1, 2, \cdots, k$). In toxicological work x_i would stand for the ith dose of bactericide, but for present purposes it stands for the midpoint of the ith interval in a frequency distribution of test scores. Of the n_i individuals falling in this interval, r_i , or a proportion $p_i = r_i/n_i$, are found to have passed the jth item of the test. It is now assumed that the p_i have independent binomial distributions, that $P_i = 1 - Q_i$ is the proportion of individuals in the population falling in the interval x_i who pass the jth item, and that P_i is a given function of two parameters α and β of the item concerned. The logistic law states that

$$P_i = [1 + \exp(-\alpha - \beta x_i)]^{-1}. \tag{4}$$

The logit of the proportion P_i, generally denoted by L_i, is defined as

$$L_i = \ln (P_i/Q_i) = \alpha + \beta x_i . \tag{5}$$

Writing $\alpha + \beta x_i = \beta(x_i - \mu)$, μ, which is equal to $-\alpha/\beta$ corresponds to $P = 0.5$, so that $-\alpha/\beta$ is the limen of the item.

For each of (2) and (4) it is seen that a transformation of the dependent variable P can be made such that if the transformed value is plotted against x a linear relationship results, the slope and intercept of which give the parameters of the original function.

The Shape of the Logistic Function

If x in (4) ranges from $-\infty$ to $+\infty$, P ranges from 0 to 1. Differentiating (4) with respect to x,

$$dP/dx = \beta P(1 - P). \tag{6}$$

Hence $dP/dx = 0$, when $P = 0$ and 1, so that the logistic curve has both a lower and an upper asymptote.

Differentiating (6),

$$d^2P/dx^2 = (1 - 2P), \tag{7}$$

which shows that the curve has a point of inflection at -0.5, halfway between the two asymptotes. The abscissa of this point, as already noted, is $-\alpha/\beta$, the limen of the item. From this description it is seen that the logistic curve resembles very closely the normal ogive. For this reason, it is, on purely empirical grounds, suitable for fitting to proportions of individuals of different levels of ability passing a test item [8].

Maximum Likelihood Estimates of the Logistic Parameters

In the ith interval, the probability that p_i is the proportion passing the item is

$$\frac{n_i\,!}{r_i\,!\,(n_i - r_i)!} P_i^{r_i} Q_i^{n_i - r_i}. \tag{8}$$

The likelihood function over all intervals is

$$\prod_i \frac{n_i\,!}{r_i\,!\,(n_i - r_i)!} P_i^{r_i} Q_i^{n_i - r_i}. \tag{9}$$

The terms in the logarithm of this function which involve the unknown parameters, α and β, are given by

$$L = \sum_i \{r_i \log_e P_i + (n_i - r_i) \log_e Q_i\}. \tag{10}$$

Substituting in (10) the value of P_i from (4) and differentiating with respect to α and β, the estimating equations for the two parameters are found to be

$$\sum n_i(p_i - \hat{P}_i) = 0, \tag{11}$$

and

$$\sum n_i x_i(p_i - \hat{P}_i) = 0, \tag{12}$$

respectively, where $\hat{P}_i = [1 + \exp(-\hat{\alpha} - \hat{\beta}x_i)]^{-1}$, $\hat{P}_i$, $\hat{\alpha}$ and $\hat{\beta}$ being estimates of the population parameters P_i, α, and β. These equations, based on the sufficient statistics $\sum n_i p_i$ and $\sum n_i p_i x_i$, have been given by Berkson [2, 4], who also shows how to solve them iteratively. The procedure is as follows. Given trial values α_0 and β_0 of the parameters required, a trial value l_0 of the logit is obtained using (5); it is $l_0 = \alpha_0 + \beta_0 x_i$, and corresponds to an estimate $\hat{p}_0$ of P_i given by

$$\hat{p}_0 = [1 + \exp(-l_0)]^{-1}.$$

Using this result, and retaining only the first term in a Taylor expansion,

$$\hat{p}_0 - \hat{P}_i \doteq (\hat{l}_0 - \hat{l}_i)\hat{P}_0\hat{Q}_0 .$$

Substituting this approximation in (11) and (12), it will be seen that the differentials $\delta\hat{\alpha}$ and $\delta\hat{\beta}$, which are the corrections to the trial values of the parameters, are given by the equations

$$\begin{aligned} \sum n_i p_i - \sum n_i \hat{p}_0 &= \delta\hat{\alpha}(\sum w_i) + \delta\hat{\beta}(\sum w_i x_i), \\ \sum n_i p_i x_i - \sum n_i \hat{p}_0 x_i &= \delta\hat{\alpha}(\sum w_i x_i) + \delta\hat{\beta}(\sum w_i x_i^2), \end{aligned} \tag{13}$$

where $w_i = n_i p_0 q_0$, corresponding to x_i. Solving these equations the corrections are found to be

$$\delta\hat{\beta} = \frac{\sum n_i p_i x_i - \sum n_i \hat{p}_0 x_i - \sum w_i x_i(\sum n_i p_i - \sum n_i \hat{p}_0)/\sum w_i}{\sum w_i x_i^2 - (\sum w_i x_i)^2/\sum w_i} \tag{14}$$

and

$$\delta\hat{\alpha} = \frac{\sum n_i p_i - \sum n_i \hat{p}_0 - \delta\hat{\beta} \sum w_i x_i}{\sum w_i}. \tag{15}$$

Anscombe [1] points out that it is unnecessary in further iterations to recalculate the coefficients on the right-hand side of equations (13), and since the expressions $\sum n_i p_i$ and $\sum n_i p_i x_i$ remain unchanged from iteration to iteration only the terms $\sum n_i \hat{p}_0$ and $\sum n_i \hat{p}_0 x_i$ have to be recalculated. "This procedure," Anscombe ([1], p. 461) remarks, "is to be compared with Finney's which is exactly modelled on the ingenious procedure due to Fisher and Bliss for fitting the integrated-normal law, for which no such sufficient statistics are available."

Estimates of the variance of $\hat{\beta}$, $\hat{\alpha}$, and $\hat{\mu}$ are given [3] by the formulas

$$\begin{aligned} \text{var } (\hat{\beta}) &= 1/\sum w_i(x_i - x)^2; \\ \text{var } (\hat{\alpha}) &= (1/\sum w_i) + \bar{x}^2 \text{ var } (\hat{\beta}); \\ \text{var } (\hat{\mu}) &= [\text{var } (\hat{\alpha}) + \text{var } (\hat{\beta})(\hat{\mu} - x)^2]/\hat{\beta}^2; \end{aligned} \tag{16}$$

where $\bar{x} = \sum w_i x_i / \sum w_i$.

A quick graphical method for estimating α and β has recently been provided by Hodges [10].

An Example

The data used in the example are taken from Ferguson's article [8] and are the same data as employed by Finney [9] in his example using probit analysis. Omitting the two extreme values of x, where the proportions may

TABLE 1

Proportions (p_i) Passing an Item at Different Levels of Ability (x_i)

x_i	n_i	p_i	$n_i p_i$
-1.94	15	0.33	4.95
-1.16	25	0.56	14.00
-0.58	43	0.70	30.10
0.00	50	0.94	47.00
0.58	43	0.95	40.85
1.16	25	0.96	24.00
1.94	15	0.93	13.95

$\sum n_i p_i = 174.85$ $\quad\quad$ $\sum n_i p_i x_i = 35.2950$

be unreliable, estimates α_0 and β_0 of the parameters were read from a rough plot of the variate x_i as abscissa and $l_i = \log_e p_i/q_i$, where $q_i = 1 - p_i$, as ordinate. These were 2.0 and 1.4 respectively, giving the equation $l_0 = 1.4\,x + 2.0$. Values of l_0 for each x-value were now calculated; these were converted to p_0-values using the table prepared by Berkson [4], and from the same table the weights $p_0 q_0$ were read. These data are tabulated in Table 2, while below the table the quantities required for substitution in (14) and (15) are recorded.

TABLE 2

Data for the First Iteration

x_i	$\hat{\ell}_o$	$\hat{p}_o$	$\hat{p}_o\hat{q}_o$	$n_i\hat{p}_o$	$w_i = n_i\hat{p}_o\hat{q}_o$
-1.94	-0.72	0.32739	0.22021	4.91085	3.30315
-1.16	0.38	0.59387	0.24119	14.84675	6.02975
-0.58	1.19	0.76674	0.17885	32.96982	7.69055
0.00	2.00	0.88080	0.10499	44.04000	5.24950
0.58	2.81	0.94321	0.05356	40.55803	2.30308
1.16	3.62	0.97392	0.02540	24.34800	0.63500
1.94	4.72	0.99116	0.00876	14.86740	0.13140

$\Sigma w_i = 25.34243$ $\quad$ $\Sigma n_i\hat{p}_o = 17654085$

$\Sigma w_i x_i = -15.535838$ $\quad$ $\Sigma n_i\hat{p}_o x_i = 34.73832$

$\Sigma w_i x_i^2 = 25.25622$ $\quad$ $\Sigma n_i n_i - \Sigma n_i\hat{p}_o = -1.69085$

$(\Sigma w_i x_i)^2 = 241.36226$ $\quad$ $\Sigma n_i p_i x_i - \Sigma n_i\hat{p}_o x_i = 0.55668$

Substituting the results from Tables 1 and 2 in (14) and (15), the first corrections are $\delta\hat{\beta} = -0.030502$ and $\delta\hat{\alpha} = -0.085419$ so that first approximations to the maximum likelihood estimates of the parameters are $\hat{\beta} = 1.3695$ and $\hat{\alpha} = 1.9146$; from these the limen of the item is found to be $-1.9146/1.3695 = -1.3980$.

For the second iteration the logit equation is $\hat{l}_o = 1.3695x + 1.9146$; one now recalculates values for the expressions $\sum n_i\hat{p}_o$ and $\sum n_i\hat{p}_o x_i$. These are 174.79230 and 35.26088, respectively, and the corrections in the second iteration are found to be $\delta\hat{\beta} = 0.004416$ and $\delta\hat{\alpha} = 0.004975$. Since these corrections have nonzero entries only after the second decimal places, further iterations are unnecessary for present purposes. The adjusted values of $\hat{\beta}$ and $\hat{\alpha}$ now are 1.3739 and 1.9196, respectively, while the new estimate of the limen of the item is -1.3972. The three successive estimates of the limen of the item, using the logistic function, are therefore -1.4286, -1.3980, and -1.3972. Convergence seems to be good.

Corresponding to the first two of these, Finney, using the probit method, obtained the values -1.32 and -1.471, respectively. While he considered his fit after one iteration to be sufficiently accurate it appears that it would have been preferable, both from the point of view of demonstrating con-

vergence and in order to get a more accurate answer, had he performed at least one further iteration.

Using (16), estimates of the standard errors of $\hat{\beta}$, $\hat{\alpha}$, and $\hat{\mu}$ are found to be

$$\text{S.E. of } \hat{\beta} = 0.0540, \quad \text{S.E. of } \hat{\alpha} = 0.2014, \quad \text{S.E. of } \hat{\mu} = 0.1956.$$

The estimate 1.3739 of β, the gradient of the logit line, is a measure of the discrimination value of the item. When compared with 0.0540, the estimate of its standard error, it is found to be highly significant.

REFERENCES

[1] Anscombe, F. J. On estimating binomial response relations. *Biometrika*, 1956, **43**, 461-464.

[2] Berkson, J. Why I prefer logits to probits. *Biometrics*, 1951, **7**, 327-339.

[3] Berkson, J. A statistically precise and relatively simple method of estimating the bio-assay with quantal response, based on the logistic function. *J. Amer. statist. Ass.*, 1953, **48**, 565-599.

[4] Berkson, J. Tables for the maximum likelihood estimate of the logistic function. *Biometrics*, 1957, **13**, 28-34.

[5] Birnbaum, A. Probability and statistics in item analysis and classification problems. Efficient design and use of tests of mental ability for various decision-making problems. Air Univ., School of Aviation Medicine, USAF, Randolph AFB, Texas, 1957.

[6] Birnbaum, A. Probability and statistics in item analysis and classification problems. Further considerations of efficiency in tests of a mental ability. Tech. Rep. No. 17, Project No. 7755-23, Air Univ., School of Aviation Medicine, USAF, Randolph AFB, Texas, 1957.

[7] Birnbaum, A. Probability and statistics in item analysis and classification problems. On the estimation of mental ability. Series Rep. No. 15, Project No. 7755-523, Air Univ., School of Aviation Medicine, USAF, Randolph AFB, Texas, 1957.

[8] Ferguson, G. A. Item selection by the constant process. *Psychometrika*, 1942, **7**, 19-29.

[9] Finney, D. J. The application of probit analysis to the results of mental tests. *Psychometrika*, 1944, **9**, 31-39.

[10] Hodges, J. L., Jr. Fitting the logistic by maximum likelihood. *Biometrics*, 1958, **14**, 453-461.

[11] Lord, F. M. An application of confidence intervals and of maximum likelihood to the estimation of an examinee's ability. *Psychometrika*, 1953, **18**, 57-76.

Manuscript received 6/19/58

Revised manuscript received 2/21/59

PSYCHOMETRIKA—VOL. 24, NO. 3
SEPTEMBER, 1959

THREE MULTIVARIATE MODELS: FACTOR ANALYSIS, LATENT STRUCTURE ANALYSIS, AND LATENT PROFILE ANALYSIS*

W. A. GIBSON†

PERSONNEL RESEARCH BRANCH, THE ADJUTANT GENERAL'S OFFICE

The factor analysis model and Lazarsfeld's latent structure scheme for analyzing dichotomous attributes are derived to show how the latter model avoids three knotty problems in factor analysis: communality estimation, rotation, and curvilinearity. Then the latent structure model is generalized into latent profile analysis for the study of interrelations among quantitative measures. Four latent profile examples are presented and discussed in terms of their limitations and the problems of latent metric and dimensionality thereby raised. The possibility of treating higher order empirical relations in a manner paralleling their various uses in the latent structure model is indicated.

At an early point in *Multiple-Factor Analysis* ([18], p. 70), Thurstone remarks:

> It would be unfortunate if some initial success with the analytical methods to be described here should lead us to commit ourselves to them with such force of habit as to disregard the development of entirely different constructs that may be indicated by improvements in measurement and by inconsistencies between theory and experiment.

This paper is an attempt to take that statement to heart.

First the derivation of the factor analysis model will be sketched, noting three inherent conceptual and procedural problems: (i) how to estimate communalities in the event that only shared variance is to be analyzed, (ii) how to resolve rotational indeterminacy, and (iii) what to do with the extra linear factors that are forced to emerge when nonlinearities occur in the data. Some recent concepts for the multidimensional analysis of qualitative data—the latent structure model of Lazarsfeld [15]—are considered, with special reference to their handling of the trouble spots in factor analysis. Next these new concepts are generalized to produce an alternative way of analyzing the interrelations among quantitative measures. This is the latent profile model.

All three of these models are discussed strictly from the point of view of sample statistics. The problem of generalizing to a population of which the

*The latter model is anticipated in an earlier paper by Green [12].

†The major portion of this paper was completed at the Center for Advanced Study in the Behavioral Sciences. The opinions expressed are those of the author and are not to be construed as reflecting official Department of the Army policy.

sample may be considered representative is not taken up for any of the three models.

Four examples of the latent profile model are given. Two of them are fictitious but their form closely parallels the corresponding two empirical examples. These examples are discussed from the viewpoint of the further work they suggest concerning the metric and dimensionality of the latent space. The use of higher order empirical relations for testing fit and for further particularization of the latent profile model is briefly discussed, and the possibility of close parallelism between developments in the latent structure and latent profile models is pointed out. The paper concludes with a plea for continued flexibility in the choice of multivariate models as new and improved ones appear.

Some Problems in Factor Analysis

The fundamental postulate of factor analysis ([18], p. 63) is expressed in the simple linear equation

$$Z_{ij} = a_{j1}Z_{i1} + a_{j2}Z_{i2} + \cdots + a_{jq}Z_{iq} . \tag{1}$$

Z_{ij} is the standard score of individual i on test j. Z_{i1} , Z_{i2} , $\cdots$, Z_{iq} are the standard scores of individual i on a hypothesized set of q statistically independent traits or factors. (This does not preclude subsequent conversion to correlated factors in any given analysis. The algebra of correlated factors will not be introduced here, however, for that would only complicate the discussion without changing the arguments.) The a's in (1) are a set of weights descriptive of test j and invariant for individuals.

Straightforward summational algebra and the independence of factors lead directly from (1) to the basic equation of factor analysis ([18], p. 78):

$$r_{jk} = a_{j1}a_{k1} + a_{j2}a_{k2} + \cdots + a_{jq}a_{kq} . \tag{2}$$

Thus r_{jk} , the correlation between tests j and k, is expressed as a simple bilinear function of the a's for those two tests. These a's, also known as factor loadings, are interpretable as correlations between tests and factors.

The essential task in the factor analysis of a battery of s tests is to solve for the a's in the system of $s(s - 1)/2$ bilinear equations resulting from (2). The number of factor loadings is sq, q for each of the s tests. These loadings can be obtained by a wide variety of techniques that have been developed over the years. Most of these methods attempt to account for the intercorrelations in terms of a minimum number of factors. Perfect accounting for the intercorrelations by the factors is seldom demanded because of sampling error and the frank expectation of at least minor disagreements between model and data. Some nonvanishing "residual" correlations are permitted, so long as they are small and show no systematic pattern.

One troublesome feature of the factor model is the problem of how to

deal with those elements in the correlation matrix that have repeated subscripts—the diagonal r_{ii}'s. To take these as perfect self-correlations of unity would amount to trying to analyze all of the variance of every test, including the unreliable part. To insert the test reliabilities into the diagonal cells would imply an interest in analyzing all of the "true" or repeatable variance of the tests, including that specific to each test and unrelated to other tests in the battery. More commonly preferred among factor analysts is to attempt to analyze only that part of a test's variance—its communality—that is shared with other tests in the battery. Communalities could be defined alternatively as those portions of the self-correlations accounted for by the factors that suffice to account for the between-test correlations. In any event, the communalities are not empirically given and yet are needed at the start for maximum efficiency of solution. In principle the communalities could be determined by certain operations (cf. [18], pp. 294–307) applied to the empirically given side entries in the correlation matrix, but these operations are usually so time consuming that a successive approximations approach is often substituted. Rough communality estimates are used to obtain an initial factorial solution, which in turn provides improved estimates for a second cycle of the same kind, and so on until the communalities are sufficiently "stabilized." It fortunately happens that with large test batteries little or no iterating of this kind is needed, but for small batteries several cycles may be required before the communality problem is adequately resolved.

A second and more important problem in factor analysis is the inherent partial indeterminacy of the a's that is known as the rotational problem. It is most easily understood in terms of q-dimensional geometry. The a's for any test j may be thought of as the projections of a point j on a set of coordinate axes in q-space. The table of factor loadings for the s tests then defines a configuration of s points in terms of their projections on a q-dimensional reference frame. But the equations of factor analysis, by themselves, do not indicate *which* position of the reference frame, among an infinite number of possibilities in the same q-space, is to be preferred. Only the origin of the coordinate system is fixed, so that the reference frame can be rotated freely from any position to any other without distortion of the configuration of points defined by it. Naturally the a's change with such rotations, but always in such a way as to preserve the spatial interrelations among the points they define. Many ways of resolving this rotational indeterminacy have been proposed. Probably the most notable is the simple structure principle ([18], pp. 181–193), which strives to simplify the factorial structure of the tests by maximizing the number of near-zero factor loadings. Many of these proposals (especially those centering around the simple structure principle) involve heavy computing loads, relatively rare geometric intuition, or both; many are debated or debatable; all are in the nature of afterthoughts that are not built into the equations defining the model.

Another perplexing problem in the factor analysis model is the paradox of difficulty factors [2, 4, 13, 20]. If factor analysis is applied to a battery of tests varying widely in difficulty but quite obviously measuring but one underlying trait, the result is not one but several "factors"—one for each level of difficulty. This is generally attributed to curvilinear relations among tests markedly different in difficulty, such curvilinearity being forced by the differential skewness of the score distributions. Note, however, that it is only the relations between tests and factors, as indicated by (1), that the factor model explicitly restricts to linear form.

The record of empirical fruitfulness (or lack thereof) of factor analysis is not at issue here. Nor are misapplications of it pertinent to this discussion. The other two models to be discussed here are meant to be put to much the same use, and they may very well suffer the same kind of misuse.

Some New Concepts: *Latent Structure Analysis*

Only one variety of latent structure analysis—that known as the discrete class model—is discussed here.

Latent structure analysis [15] is Lazarsfeld's technique for analyzing the interrelations among dichotomous attributes, such as the item responses on a survey questionnaire. It is based on linear recruitment equations of the following kind:

$$
\begin{aligned}
n &= n_1 + n_2 + \cdots + n_q \,, \\
n_j &= n_1 p_{1j} + n_2 p_{2j} + \cdots + n_q p_{qj} \,, \\
n_{jk} &= n_1 p_{1jk} + n_2 p_{2jk} + \cdots + n_q p_{qjk} \,, \\
n_{jkl} &= n_1 p_{1jkl} + n_2 p_{2jkl} + \cdots + n_q p_{qjkl} \,, \quad \text{etc.}
\end{aligned} \tag{3}
$$

The quantities on the left are empirically given or *manifest* data. They indicate the number of people in the entire sample, n, the number endorsing a single item, n_j , the number endorsing any pair of items, n_{jk} , and so on. The quantities on the right are the underlying or *latent* parameters of the model. The number of terms on the right is q, the number of mutually exclusive and exhaustive subgroups (*latent classes*) into which the analysis will divide the total sample. The number of people in latent class 1 is n_1 , and so on. The *latent probability* p_{1j} is the proportion of the members of latent class 1 who endorse item j, p_{1jk} is the proportion of class 1 members who endorse both items j and k, and so on. Equations (3) merely show how the manifest joint endorsements are recruited from the latent classes. It may be noted, in passing, that equations (3), being a set of recruitment equations, are intrinsically linear, while the initial equation of factor analysis is linear only because it was made so.

The manifest-latent distinction that is so prominent in latent structure theory is of course central in factor analysis as well. There the tests and their intercorrelations are manifest, and the factors and the loadings thereon are latent. The same distinction also appears in the latent profile model to be discussed later.

The preceding recruitment equations are equally valid for any method of classification and for any number of latent classes. The next step is to invoke a pertinent basis for classification. This is the core of the latent structure model. It is quite relevant to require that each latent class exhibit homogeneity with respect to any underlying dimensions that may be responsible for the manifest interrelations. Perfect homogeneity is not crucial, so long as deviations from the class norm are random. Such random deviations are of course uncorrelated within the class. Thus it is sufficient to require that each latent class be homogeneous enough, with respect to any and all such latent dimensions, so that all item responses *within the class* are independent in the coin-tossing sense. This intraclass independence is expressed in the following equations.

$$
\begin{aligned}
(4)\qquad & p_{1jk} = p_{1j}p_{1k}\,, \qquad p_{2jk} = p_{2j}p_{2k}\,, \cdots, p_{qjk} = p_{qj}p_{qk}\,, \\
& p_{1jkl} = p_{1j}p_{1k}p_{1l}\,, \qquad p_{2jkl} = p_{2j}p_{2k}p_{2l}\,, \cdots, p_{qjkl} = p_{qj}p_{qk}p_{ql}\,, \quad \text{etc.}
\end{aligned}
$$

The substitution of (4) into (3) yields the basic equations of latent structure analysis ([15], p. 385).

$$
\begin{aligned}
& n = n_1 + n_2 + \cdots + n_q\,, \\
(5)\qquad & n_j = n_1p_{1j} + n_2p_{2j} + \cdots + n_qp_{qj}\,, \\
& n_{jk} = n_1p_{1j}p_{1k} + n_2p_{2j}p_{2k} + \cdots + n_qp_{qj}p_{qk}\,, \\
& n_{jkl} = n_1p_{1j}p_{1k}p_{1l} + n_2p_{2j}p_{2k}p_{2l} + \cdots + n_qp_{qj}p_{qk}p_{ql}\,, \quad \text{etc.}
\end{aligned}
$$

Thus all of the manifest joint frequencies are accounted for in terms of $(q + sq)$ latent parameters, q class sizes and q latent probabilities $(p_{1j}\,, p_{2j}\,, \cdots, p_{qj})$ for each of the s items. The successive levels of manifest frequencies $(n, n_j\,, n_{jk}\,,$ etc.) number, respectively, 1, s, $s(s-1)/2$, etc., the coefficients in the binomial expansion $(a + b)^s$. These add up to 2^s, the number of equations relating manifest to latent data in this model.

The task here, as in factor analysis, is to solve the basic equations for the unknown latent parameters. Several latent structure solutions already exist. The most recent of these [1, 7] avoid the use of any joint frequencies with repeated subscripts $(n_{jj}\,, n_{jjk}\,, n_{jjj}\,, n_{jjkl}\,,$ etc.). In latent structure analysis these are treated as analogous to the communalities of factor analysis, in not being manifest. To interpret them as equivalent to the corresponding

lower order joint frequencies without repeated subscripts (i.e., $n_{ij} = n_i$, $n_{ijk} = n_{ik}$, $n_{iij} = n_i$, etc.) would be analogous to the use of unit self-correlations in factor analysis. Instead, the latent structure model usually treats these elements with repeated subscripts as stemming from, rather than leading to, the latent parameters of the model that suffices to account for the manifest data *without* repeated subscripts.

The Anderson latent structure solution and an earlier one by Green [11] eliminate the latent structure analogue of the rotational problem through the use of manifest data with more than two subscripts, such as n_{ikl}. (A cursory investigation indicates that it might be self-contradictory for the factor model to make use of manifest interrelations among more than two variables at a time. This is because at least some of the joint distributions among factor scores would have to exhibit asymmetries that could easily destroy the basic linear postulate of factor analysis.)

These higher order data select, from the infinite number of rotational solutions accounting for lower order manifest frequencies equally well, the *one* that fits *themselves* best. Another early latent structure solution [5, 6, or 8] effects a partial (often severe) reduction of rotational indeterminacy without the expense of obtaining higher order data. This is accomplished by capitalizing on the simple fact that the latent parameters, being probabilities, can be neither negative nor greater than unity.

Earlier it was indicated that the artifact of difficulty factors arises in factor analysis when curvilinearities are present. Note that the derivation of the latent structure model embodies no restriction on curvilinear relations, either among the manifest attributes or between them and the latent dimensions.

The latent structure model has already shown promise in empirical research [cf. 14, 17]. Only its restriction to the analysis of dichotomous attributes prevents it from being a feasible alternative to the factor analysis of quantitative variables. The latent profile model to be taken up next is the generalization of latent structure analysis to the case of quantitative manifest variables [cf. 12].

Some Linear Recruitment Equations for Quantitative Manifest Variables

A reasonable question to ask is whether the natural linearity of recruitment equations could not provide a basis for analyzing interrelations among manifest variables that are quantitative rather than qualitative. This section will display such a system of recruitment equations.

Suppose there is available a set of s quantitative measures, such as test scores, on a sample of n people. On some basis let every sample member be assigned to one, and only one, of q subgroups. Then the size, sums of scores, and sums of score products for the entire sample are related to corresponding subgroup statistics in the following simple way.

$$
\begin{aligned}
n &= n_1 + n_2 + \cdots + n_q \,, \\
\sum^{n} X_{ij} &= \sum^{n_1} X_{ij} + \sum^{n_2} X_{ij} + \cdots + \sum^{n_q} X_{ij} \,, \\
\sum^{n} X_{ij}X_{ik} &= \sum^{n_1} X_{ij}X_{ik} + \sum^{n_2} X_{ij}X_{ik} + \cdots + \sum^{n_q} X_{ij}X_{ik} \,, \\
\sum^{n} X_{ij}X_{ik}X_{il} &= \sum^{n_1} X_{ij}X_{ik}X_{il} \\
&\quad + \sum^{n_2} X_{ij}X_{ik}X_{il} + \cdots + \sum^{n_q} X_{ij}X_{ik}X_{il} \,, \quad \text{etc.}
\end{aligned} \tag{6}
$$

All summations in (6) are over individuals. The summations on the left are over the entire sample, while those on the right are over the members of the various subgroups. X_{ij} is the score of individual i on test j, and it may be in raw, deviational, standard, or any other derived units. The same is true of X_{ik} , X_{il} , etc.

It is convenient to designate the various summations in (6) by the letter m with appropriate subscripts. Thus m_j and m_{jk} represent, respectively, the sum of scores on test j and the sum of products of scores on tests j and k, each sum being for the entire sample. m_{1j} and m_{1jk} , on the other hand, stand for the same things in subgroup 1 only, and so on. Equations (6) then become

$$
\begin{aligned}
n &= n_1 + n_2 + \cdots + n_q \,, \\
m_j &= m_{1j} + m_{2j} + \cdots + m_{qj} \,, \\
m_{jk} &= m_{1jk} + m_{2jk} + \cdots + m_{qjk} \,, \\
m_{jkl} &= m_{1jkl} + m_{2jkl} + \cdots + m_{qjkl} \,, \quad \text{etc.}
\end{aligned} \tag{7}
$$

The m's in (7) will be referred to as *product moments* of various orders. The product moments in the third line, involving two tests, will be said to be of second order, while those for three tests are of third order, and so on. For consistency, the m's involving only one test j will be called first-order product moments, and n, the sample size, could be called the zero-order product moment for the sample. The m's on the left in (7) will be called sample product moments, while those on the right are subgroup product moments.

It should be stressed that (6) and (7) in no way restrict either the empirical data or the method of classification into subgroups. It will be the burden of the next two sections to establish a basis for grouping that will convert these recruitment equations into a mathematical model.

The Fundamental Theorem of Latent Profile Analysis

Consider now a two-dimensional joint distribution in which the score of every person in the sample on test k is plotted against his score on test j.

Such a scatterplot is pictured in Figure 1. The unit of measurement for both tests is entirely arbitrary here. The ellipse in Figure 1 indicates only one of the possible shapes that the total configuration of points could have. The circle represents a subgroup t, of size n_t, *within* which the correlation between

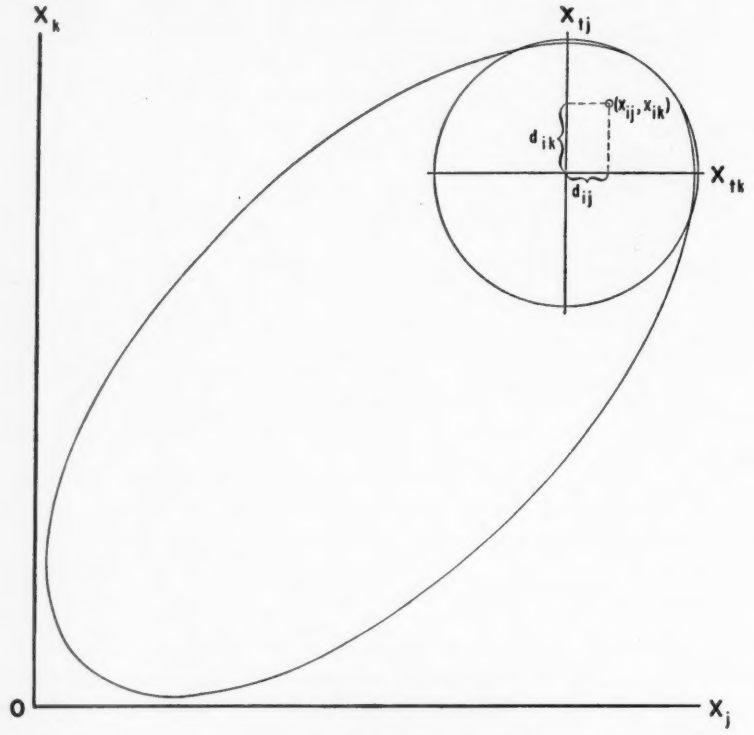

FIGURE 1
A Hypothetical Scatter Diagram

tests j and k is zero. The two lines labeled X_{tj} and X_{tk} in Figure 1 indicate the means of subgroup t on tests j and k. They intersect in what is called the *centroid* of the points comprising subgroup t. The point corresponding to individual i, a member of subgroup t, is shown with the coordinates (X_{ij}, X_{ik}), his scores on the two tests. The two distances, d_{ij} and d_{ik}, are the deviations of individual i from the j and k means of his subgroup. A property of such deviations is that their sum, over all members of the subgroup, is zero.

The j and k scores of individual i may be expressed in terms of his subgroup means and his deviations from those means, as follows:

$$X_{ij} = X_{tj} + d_{ij} , \tag{8}$$

$$X_{ik} = X_{tk} + d_{ik} . \tag{9}$$

Then m_{tj} , the sum of j scores for subgroup t, is given by

$$\begin{aligned} m_{tj} &= \sum^{n_t} X_{ij} \\ &= \sum^{n_t} (X_{tj} + d_{ij}) \\ &= \sum^{n_t} X_{tj} + \sum^{n_t} d_{ij} \\ &= n_t X_{tj} . \end{aligned} \tag{10}$$

The first term in the third line simplifies because X_{tj} is the same for every member of the subgroup. The last term in the third line vanishes because it is a sum of deviation scores. Thus the j scores for any subgroup t contribute to the sum of all j scores *as if* the members of the subgroup were concentrated at their mean for test j. This result (by no means a new one) will next be generalized, in an appropriate way, to the second-order product moment, m_{tjk} , for subgroup t.

By definition and with the help of (8) and (9), m_{tjk} becomes

$$\begin{aligned} m_{tjk} &= \sum^{n_t} X_{ij} X_{ik} \\ &= \sum^{n_t} (X_{tj} + d_{ij})(X_{tk} + d_{ik}) \\ &= \sum^{n_t} (X_{tj} X_{tk} + X_{tj}\, d_{ik} + X_{tk}\, d_{ij} + d_{ij}\, d_{ik}) \\ &= \sum^{n_t} X_{tj} X_{tk} + X_{tj} \sum^{n_t} d_{ik} + X_{tk} \sum^{n_t} d_{ij} + \sum^{n_t} d_{ij}\, d_{ik} \\ &= n_t X_{tj} X_{tk} . \end{aligned} \tag{11}$$

The first term in the fourth line of (11) simplifies because X_{tj} and X_{tk} are the same for all members of the subgroup. The second and third terms in that line vanish because they contain sums of deviations. The last term vanishes because it is the numerator of the formula for the correlation, within the subgroup, between tests j and k. The subgroup was earlier defined as having that correlation equal to zero. Thus the second-order product moment, m_{tjk} , is the same as it would be if all members of subgroup t were concentrated at their centroid. This holds for any subgroup within which tests j and k are uncorrelated.

A further distinction in terminology needs to be made here. The quantity

m_{tjk} is a sum of products of horizontal and vertical distances from the origin in Figure 1. It will therefore be called a *product moment about the origin.* Another kind of product moment in (11) is the last term in line 4, the sum of products of horizontal and vertical distances of a set of points from their own centroid. Such a quantity is appropriately called a *product moment about the centroid* of the set of points that is involved. The geometric interpretation of scores and deviations as distances makes it clear that the point of reference (origin, centroid, or some other point) of a product moment must always be specified in some way. Naturally this holds for all orders of product moments. Up to now, with four exceptions, all product moments have had the origin as their point of reference. The four exceptions are the last term in line 3 of (10) and the summations in the last three terms of line 4 in (11), which have the centroid of subgroup t as their point of reference.

The results of (10) and (11) can be generalized to higher order product moments by imposing additional restrictions on subgroup t. Not only must that subgroup be defined as having all pairs of tests j, k, and l uncorrelated within it, but, let it also have, for those three tests, a vanishing third-order product moment about its centroid. With these restrictions its third-order product moment about the origin, m_{tjkl} , becomes

$$m_{tjkl} = n_t X_{tj} X_{tk} X_{tl} \,. \tag{12}$$

This result is, in form and mode of development, a third-order analogue of the final step in the previous two equations. The fourth-order equivalent is obtained by analogous higher order restrictions, and so on.

Since in all of this discussion the origin could have been placed at any point 0, it now becomes possible to state the fundamental theorem of latent profile analysis.

> The g-order product moment, about any point O, of n_t points having zero product moments of order g and less about their centroid, is equal to the g-order product moment, about the point O, of n_t points placed at that centroid.

The Basic Equations of Latent Profile Analysis

The foregoing theorem provides a basis for grouping in the recruitment equations introduced previously. The close analogy with latent structure analysis will be obvious. Each subgroup or latent class should be homogeneous in whatever underlying dimensions are necessary to account for the observed interrelations. The homogeneity need not be complete, so long as deviations from the class averages are random, i.e., independent.

In the statistics of dichotomous attributes (as employed in coin-tossing experiments, for example), the notion of independence has usually applied to all orders of joint occurrence, and not just to pairs of events. This is the case in latent structure analysis, where within-class independence is *defined*

as pertaining not only to all pairs of items but also to all triplets, all quadruplets, and so on. (It is easily shown by example that higher order independence among dichotomous attributes is *not* a mere logical consequence of uncorrelatedness between all pairs of such attributes.)

The concept of uncorrelatedness among quantitative measures, on the other hand, has more often been restricted, at least among psychometricians, to pairs of such measures. (A statistically oriented prepublication reviewer has pointed out that the two kinds of independence being discussed here are known to statisticians as *pair-wise* and *mutual* independence.) This is not an intrinsic or logical difference between qualitative and quantitative statistics. It is rather only a historical accident that the question of higher order independence among quantitative measures has less often arisen in psychometrics. That question arises here, for it turns out that the proper definition of such independence is crucial for this model.

In the previous section the within-class uncorrelatedness between pairs of tests was shown to be synonymous with vanishing second-order product moments about the centroid of the class. This is because such product moments are the numerators of the formulas for the corresponding correlations. Purely by analogy, higher order within-class independence may be equated with the vanishing of higher order product moments about the centroid of the class. (The failure of such product moments to vanish would allow, for example, a positive correlation between tests j and k among class members with high scores on test l, accompanied by a compensating negative correlation between the same two tests among class members having low scores on test l. This could happen in spite of zero correlations between all pairs of the three tests within the class as a whole. If correlational patterns can differ within subdivisions of a class, then the class is not homogeneous even from a commonsense point of view.) Therefore let the within-class independence of the present model be *defined* as applying to all orders of interrelations, and as expressing itself in the vanishing of product moments of all orders about the centroid of the class. Then the fundamental theorem applies with full force to the product moments of each class, so that the results of equations (10), (11), and (12) can be used to transform (7) into the basic equations of latent profile analysis:

$$
\begin{aligned}
n &= n_1 + n_2 + \cdots + n_q \,, \\
m_j &= n_1 X_{1j} + n_2 X_{2j} + \cdots + n_q X_{qj} \,, \\
m_{jk} &= n_1 X_{1j} X_{1k} + n_2 X_{2j} X_{2k} + \cdots + n_q X_{qj} X_{qk} \,, \\
m_{jkl} &= n_1 X_{1j} X_{1k} X_{1l} + n_2 X_{2j} X_{2k} X_{2l} + \cdots + n_q X_{qj} X_{qk} X_{ql} \,, \quad \text{etc.}
\end{aligned} \tag{13}
$$

Thus for s tests the 2^s manifest product moments (including n) are accounted for in terms of $(q + sq)$ latent parameters—the q latent class sizes and the

q class averages for each of the s tests. Each latent class is therefore characterized by its size and its profile of s test averages, its *latent profile.*

The latent profile equations and those of latent structure analysis turn out to be identical in form, as can be seen from a comparison of (5) and (13). This means that all algebraic latent structure solutions [1, 7, 11] are directly applicable to the latent profile equations. Hence the latent profile equations have a solution that can be obtained without the involvement of communality analogues (m_{ij} , m_{ijk} , m_{iji} , etc.), and that is, in general, rotationally unique. (A conversation with Robert P. Abelson at Yale University has clarified the fact that, when deviational scores are used, the m_{ij} are the between-groups variances.) Nor do the latent profile equations restrict the occurrence of curvilinear relations among tests or between tests and underlying dimensions. Thus the dilemma of difficulty factors is avoided in this model.

Two Special Cases of Latent Profile Analysis

In the development of the latent profile equations the score units were entirely arbitrary. There are, however, two kinds of test scores that deserve special attention. Consider first the case where the manifest variables are, as in latent structure analysis, dichotomous attributes. Let the presence and absence of each such attribute be designated by scores of one and zero, respectively. In this case the manifest latent profile m's become identical with the manifest latent structure n's, and the class averages of latent profile analysis become the latent probabilities of latent structure analysis. The latent class sizes mean the same thing in both models. Thus the latent structure model is interpretable as the special case of latent profile analysis in which the manifest variables are dichotomous.

A second special result is obtained by using standard scores and by dividing the latent profile equations through by n, the number of people in the sample. The latent profile equations then assume their standard form:

$$
\begin{aligned}
1 &= p_1 + p_2 + \cdots + p_q \ , \\
0 &= p_1 Z_{1j} + p_2 Z_{2j} + \cdots + p_q Z_{qj} \ , \\
r_{jk} &= p_1 Z_{1j} Z_{1k} + p_2 Z_{2j} Z_{2k} + \cdots + p_q Z_{qj} Z_{qk} \ , \\
r_{jkl} &= p_1 Z_{1j} Z_{1k} Z_{1l} + p_2 Z_{2j} Z_{2k} Z_{2l} + \cdots + p_q Z_{qj} Z_{qk} Z_{ql} \ , \quad \text{etc.}
\end{aligned}
\tag{14}
$$

The p's are the proportionate class sizes, and of course their sum is unity. The Z's are the average standard scores of classes on tests. Their weighted sum for any test (in the second line) vanishes because it is the mean of all standard scores on that test. The r_{jk} , being average products of pairs of standard scores, are the same correlations for which factor analysis attempts to account. The r_{jkl} are, analogously, average triple products, and so on.

The latent profile equations in standard form have the advantage of dealing with magnitudes that are independent of sample size and of arbitrary score units. It is in standard form that the equations will be applied to the examples in the next four sections.

Latent Profile Example I: *A Fictitious Two-Class Case*

The fictitious manifest data in Tables 1 and 2 will serve as a first latent profile example. Imagine the data as resulting from the administration of four alternate forms of the same test (such as arithmetic) to a sample of, let us say, a hundred people. Suppose, further, that all six intercorrelations turn out to be .50, so that every two-dimensional scatter diagram, with scores plotted in standard units, has the appearance of an ellipse twice as long as wide, centered on the origin, and tilted at an angle of 45 degrees. There are four three-dimensional scatter diagrams. Each is approximately egg-shaped, and, being symmetric about the origin, yields a third-order manifest product moment of zero.

Tables 1 and 2 display the necessary manifest data in a convenient way. The upper left entry in Table 1 is the first term in the first line of (14), the latent profile equations in standard form. The other entries in row and column 0 of Table 1 are the means of standard scores for each of the four tests—the left-hand term in the second line of (14). The remaining cells of Table 1 contain the test intercorrelations—the manifest data in the third line of (14).

Table 2 summarizes the manifest data of first, second, and third order. The upper left entry is the sum of the four means of standard scores. Each of the other entries in row or column 0 is the sum of the four correlations (including r_{ii}) involving the associated test. Every other cell in Table 2 contains the sum of the four third-order manifest product moments (including r_{iik} and r_{ikk}) for the corresponding pair of tests.

For convenience of exposition in both fictitious examples in this paper, all elements with repeated subscripts (such as r_{ij} , r_{iik} , and r_{iii}) are treated as known. Their values are, in fact, easily inferred from the simple form of the manifest data, but this would not be true generally, even for all sets of fictitious data. In this first example, all r_{ij} are .50, all r_{iik} are zero, and all r_{iii} are zero.

Tables 1 and 2 have been labeled R and R_1 respectively. This is convenient notation for any such display of manifest data, and it will be used in all examples. In any latent profile solution, a distinction must be made between the given and the fitted R and R_1 , the latter pair of tables indicating what the former should be in order to be completely accounted for by the solution. In both fictitious examples in this paper, the given and the fitted manifest data, the latter computed from the solution by means of (14), are identical and hence need not be compared. In the two empirical examples,

TABLE 1

R for a Fictitious Two-Class Case

Test No.	Test Number 0	1	2	3	4
0	1.00	.00	.00	.00	.00
1	.00	.50	.50	.50	.50
2	.00	.50	.50	.50	.50
3	.00	.50	.50	.50	.50
4	.00	.50	.50	.50	.50

TABLE 2

R_1 for a Fictitious Two-Class Case

Test No.	Test Number 0	1	2	3	4
0	.00	2.00	2.00	2.00	2.00
1	2.00	.00	.00	.00	.00
2	2.00	.00	.00	.00	.00
3	2.00	.00	.00	.00	.00
4	2.00	.00	.00	.00	.00

TABLE 3

Latent Profile Solution for a Fictitious Two-Class Case

	Test No.	Latent Class I	II
Class Means	1	-.71	.71
	2	-.71	.71
	3	-.71	.71
	4	-.71	.71
Class Sizes		.50	.50

TABLE 4

Given Correlations for Nine Reading Tests

Test No.	Test Number 0	1	2	3	4	5	6	7	8	9
0	1.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
1	.00	...	.72	.41	.28	.52	.71	.68	.51	.68
2	.00	.72	...	.34	.36	.53	.71	.68	.52	.68
3	.00	.41	.34	...	.16	.34	.43	.42	.28	.41
4	.00	.28	.36	.16	...	.30	.36	.35	.29	.36
5	.00	.52	.53	.34	.30	...	.64	.55	.45	.55
6	.00	.71	.71	.43	.36	.64	...	.76	.57	.76
7	.00	.68	.68	.42	.35	.55	.76	...	.59	.68
8	.00	.51	.52	.28	.29	.45	.57	.59	...	.58
9	.00	.68	.68	.41	.36	.55	.76	.68	.58	...

TABLE 5

Approximate Latent Profile Solution for Nine Reading Tests

	Test No.	Latent Class I	II
Class Means	1	-.80	.80
	2	-.81	.81
	3	-.47	.47
	4	-.41	.41
	5	-.68	.68
	6	-.90	.90
	7	-.85	.85
	8	-.66	.66
	9	-.84	.84
Class Sizes		.50	.50

however, the comparison will be made when possible in order to appraise the adequacy of the solution.

The latent profile solution for the present example, obtained by applying the algebra of the latent structure solution of Green [11], is shown in Table 3. Each of the two latent classes is defined in terms of its relative size and its latent profile—the complete set of average test scores for its members. Apparently Class I consists of those who are poor at arithmetic, while Class II contains the good arithmetic students.

A fruitful way to visualize this latent profile solution is in terms of the regressions of the tests on the latent continuum of arithmetic ability. Such a graph of mean test score, Y, against position along the latent continuum, X, is shown in Figure 2. Here the regressions of all four tests on the latent

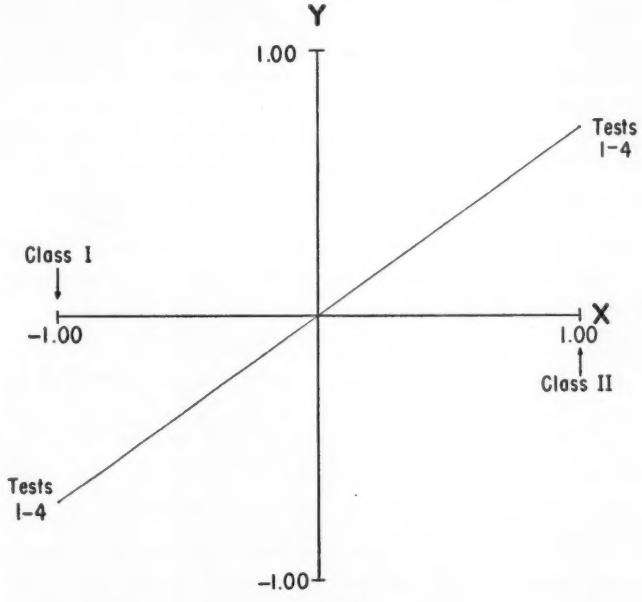

FIGURE 2
Regression of Tests on Latent Continuum for A Fictitious Two-Class Case

continuum are identical. In Figure 2 both Y and X are expressed in standard units, so that the slope of the regression line is also the correlation between Y and X. This is a correlation between test and "factor," and for the linear regressions of the present example, these correlations turn out to be exactly

the same ($\sqrt{2}/2$) as the factor loadings that would result from a factor analysis of the correlations in Table 1. This simple correspondence between the two models will vanish, however, as soon as any of the regressions become nonlinear.

Latent Profile Example II: *An Empirical Two-Class Case*

As a second latent profile example, consider the given R in Table 4. That table is a simple modification of a table of intercorrelations among nine reading tests previously reported by Davis [3]. The only modification was to border Davis' table with the 0 column and row. The intercorrelations were based on 421 cases.

A letter from Davis has indicated that the raw scores that would be needed for the computation of R_1 are no longer available. In the absence of higher order data which would provide a unique solution, it was necessary here to adapt some factoring and rotational procedures that were involved in an early approximate latent structure solution [5, 6, or 8] in order to obtain an approximate latent profile solution. For this purpose a factorization of the Davis data by Thurstone [19] was used. Thurstone's analysis indicated that one factor was sufficient to account for the data, all but three of the discrepancies between given and fitted correlations being less than .04, and the largest being .07. It is with exactly these same discrepancies that the present latent profile solution accounts for the intercorrelations.

An approximate latent profile solution for the Davis data is shown in Table 5. This solution is obtainable by resolving the rotational problem with any one of the following three assumptions:

(i) that the two latent classes are equal in size;

(ii) that the two latent profiles are identical except for reversed algebraic sign; or

(iii) that the given R_1 , if available, would be like that of Example I in having nonzero entries only in its 0 column and row.

The solution in Table 5 generates, by means of (14), a fitted R_1 having the form indicated in assumption 3.

The regressions for this solution are pictured in Figure 3. Again both axes are in standard units, so that the slopes of the regressions are identical with the factor loadings reported by Thurstone. Although the present latent profile solution is not rotationally unique, it can fairly readily be shown to possess an important kind of invariance, namely, that *the slopes of the regressions, when both axes are in standard units, remain constant regardless of how the rotational indeterminacy is resolved.* This is but one illustration of the fact that the factor analysis and latent profile models are mutually complementary when the assumptions of both models are not violated.

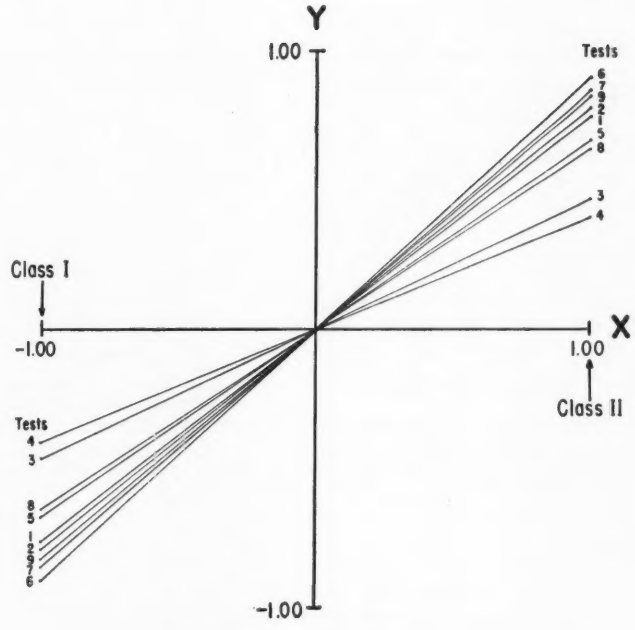

FIGURE 3
Regressions of Tests on Latent Continuum for Nine Reading Tests

Latent Profile Example III: *A Fictitious Three-Class Case*

The fictitious R and R_1 in Tables 6 and 7 will provide a first illustration of how latent profile analysis handles the problem of difficulty factors. Imagine tests 1 and 2 as being two easy vocabulary tests, tests 4 and 5 as two hard vocabulary tests, and 3 as a vocabulary test of intermediate difficulty. Again assume the data are based upon a hundred cases.

Before proceeding to the latent profile analysis of this fictitious data, it will be instructive to examine the results of a factor analysis of the correlations in Table 6. The simple structure factor analytic solution with correlated factors is given in Table 8. The entries in that table are the correlations between the five tests and the two factors, A and B. The correlation between the two factors, r_{AB}, is .33. If the usual rules for interpreting factors were followed unquestioningly here, the conclusion would be that the two factors are knowledge of easy words, A, and knowledge of hard words, B, and that the two abilities are relatively independent. This is absurd.

The unique, perfectly fitting latent profile solution for this example, again obtained from R and R_1 by the same algebra as is used in the latent

TABLE 6

R for a Fictitious Three-Class Case

Test No.	Test Number 0	1	2	3	4	5
0	1.00	.00	.00	.00	.00	.00
1	.00	.75	.75	.50	.25	.25
2	.00	.75	.75	.50	.25	.25
3	.00	.50	.50	.50	.50	.50
4	.00	.25	.25	.50	.75	.75
5	.00	.25	.25	.50	.75	.75

TABLE 7

R_1 for a Fictitious Three-Class Case

Test No.	Test Number 0	1	2	3	4	5
0	.00	2.50	2.50	2.50	2.50	2.50
1	2.50	-2.50	-2.50	-1.25	.00	.00
2	2.50	-2.50	-2.50	-1.25	.00	.00
3	2.50	-1.25	-1.25	.00	1.25	1.25
4	2.50	.00	.00	1.25	2.50	2.50
5	2.50	.00	.00	1.25	2.50	2.50

TABLE 8

Simple Structure Factor Analysis Solution for Correlations in Table 6

Test No.	Factors A	B
1	.82	.00
2	.82	.00
3	.41	.41
4	.00	.82
5	.00	.82

$r_{AB} = .33$

TABLE 9

Latent Profile Solution for a Fictitious Three-Class Case

	Test No.	Latent Class I	II	III
Class Means	1	-1.50	.50	.50
	2	-1.50	.50	.50
	3	-1.00	.00	1.00
	4	- .50	-.50	1.50
	5	- .50	-.50	1.50
Class Sizes		.25	.50	.25

structure solution of Green [11], is shown in Table 9. Figure 4 shows, in standard units, the regressions implied by Table 9 and by the assumption of equal spacing of the latent classes along the single latent continuum of vocabulary knowledge.

The contour of the various regressions in Figure 4 is exactly what would be expected on the basis of the relative difficulty of the tests. The easy tests (1 and 2) discriminate only at the lower end of the latent continuum. The hard tests (4 and 5) differentiate only at the upper end. Test 3, of medium difficulty, discriminates throughout the range.

Latent Profile Example IV: *An Empirical Three-Class Case*

The data in Table 10 will provide a final latent profile example. Aside from its 0 column and row, that table is merely a rounded version of a table of subtest intercorrelations reported by Ferguson ([4], p. 328) to illustrate the occurrence of difficulty factors. Ferguson took single items from a Moray-House verbal intelligence test and combined them into six subtests that were reasonably homogeneous in content but that increased in difficulty from subtest 1 to subtest 6. The correlations were based on a sample of 108 children, age 11.

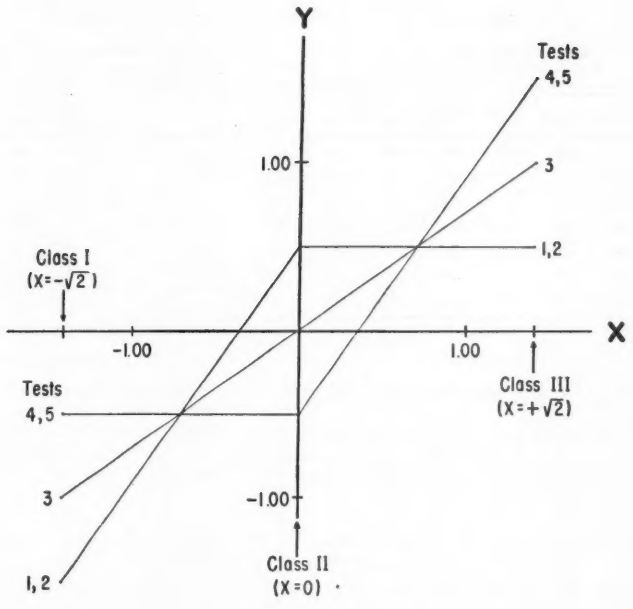

FIGURE 4
Regressions of Tests on Latent Continuum for A Fictitious Three-Class Case

The simple structure correlated factor solution for the Ferguson example is shown in Table 12. That solution was obtained by a rotation of the factorization reported by Ferguson ([4], p. 328). Here again it would be absurd to conclude, according to the usual rules, that the data must be thought of in terms of two relatively independent factors—high-level and low-level verbal intelligence.

A letter from Ferguson has indicated that the raw scores are no longer available for the computation of higher order manifest product moments. It therefore was necessary to approximate the latent profile solution by means of procedures similar to those used in the Davis example. For this purpose Ferguson's factorization of his correlations was used. His two factors accounted for the correlations with discrepancies not exceeding .03, and the present latent profile solution fits the correlations in eactly the same way.

The rotational indeterminacy here was resolved in two stages. The first step was to rotate Ferguson's factorization, with attention being given only to subtests 1 and 6, into maximum correspondence with the initial factorization for the first and last tests in Example III. The latter factorization was a part of the latent profile solution for that example. In Example III the unique solution was obtained by applying, to the initial factorization

TABLE 10

Given Correlations for Six Verbal Intelligence Sub-Tests

Test No.	Test Number 0	1	2	3	4	5	6
0	1.00	.00	.00	.00	.00	.00	.00
1	.00	...	.86	.81	.81	.61	.37
2	.00	.86	...	.80	.82	.68	.47
3	.00	.81	.80	...	.87	.80	.67
4	.00	.81	.82	.87	...	.80	.68
5	.00	.61	.68	.80	.80	...	.78
6	.00	.37	.47	.67	.68	.78	...

TABLE 11

Fitted R_1 for Six Verbal Intelligence Sub-Tests

Test No.	Test Number 0	1	2	3	4	5	6
0	.00	4.31	4.49	4.80	4.84	4.47	3.77
1	4.31	-4.42	-3.93	-2.98	-3.07	-1.74	- .24
2	4.49	-3.93	-3.37	-2.30	-2.39	-1.02	.45
3	4.80	-2.98	-2.30	- .99	-1.08	.33	1.75
4	4.84	-3.07	-2.39	-1.08	-1.16	.26	1.70
5	4.47	-1.74	-1.02	.33	.26	1.52	2.71
6	3.77	- .24	.45	1.75	1.70	2.71	3.57

TABLE 12

Simple Structure Factor Analysis Solution for Correlations in Table 10

Test No.	Factors A	B
1	.86	.00
2	.75	.15
3	.55	.43
4	.57	.42
5	.30	.64
6	.00	.81

$r_{AB} = .43$

TABLE 13

Approximate Latent Profile Solution for Six Verbal Intelligence Sub-Tests

	Test No.	Latent Class I	II	III
Class Means	1	-1.64	.51	.62
	2	-1.56	.36	.84
	3	-1.38	.07	1.24
	4	-1.41	.09	1.23
	5	-1.05	-.20	1.44
	6	- .60	-.48	1.56
Class Sizes		.25	.50	.25

of R, a rotation completely specified by the higher order data in R_1 . The second rotational stage in the present approximate solution was to imitate the Example III solution by using exactly the same rotation. The consequences of this rotational solution are the following.

(i) The relative class sizes are the same as in Example III.

(ii) The regressions of all six subtests on the latent continuum are ascending.

(iii) Assuming equal spacing of the classes along the latent continuum, the regressions of subtests 1 and 6 have curvatures that are approximately equal but opposite in direction.

(iv) The form of the fitted R_1 is similar to that of the given and fitted R_1 of Example III.

The resulting approximate latent profile solution is given in Table 13, and the regressions, again in standard units, and assuming equal spacing of the three classes along the latent continuum, are shown in Figure 5. The progression of curvatures from the easiest to the hardest subtest is just what it ought to be. This progression was found to be quite invariant over a wide range of alternative approximate solutions. Several such solutions were

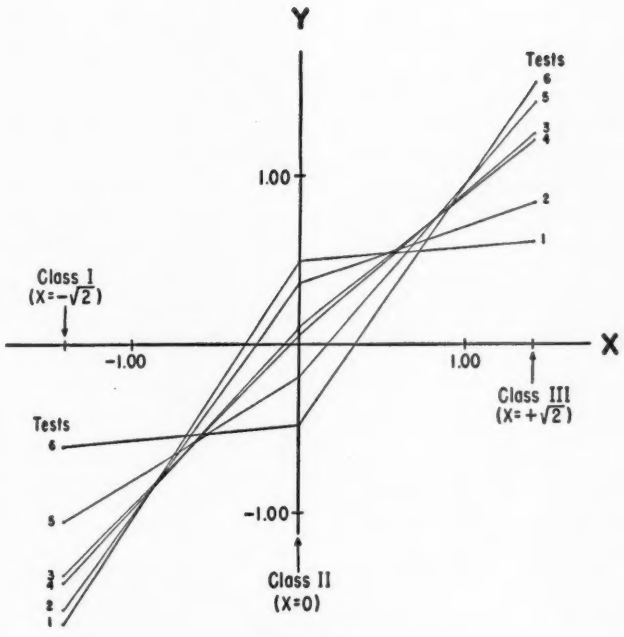

FIGURE 5
Regressions of Subtests on Latent
Continuum for Six Verbal Intelligence Subtests

computed in order to study the nonuniqueness of this latent profile solution. Only one restriction applied to all of the alternative solutions that were tried, namely, that the resulting regressions be ascending for all subtests. Within this restriction large changes in class sizes and in class averages could be brought about, but never in such a way as to alter the ordering of curvatures among the regressions. It should be added that only with very strained assumptions about the spacing of the three classes along the latent continuum could the regressions of both subtests 1 and 6 be made to curve in the same direction.

Table 11 shows the fitted R_1 implied by the approximate solution in Table 13. A comparison of Tables 7 and 11 will reveal the similarities between the fitted higher order manifest data for the two three-class latent profile examples.

Discussion

There are two limitations to the foregoing latent profile examples that should be discussed explicitly. These are in addition to the indeterminacy

produced by the absence of higher order manifest data in the case of the two empirical examples. The first is the lack of a scale of measurement for the latent continuum. For two classes this is unimportant, for it leaves only an arbitrariness as to the origin and the distances of the two classes from that origin. The problem of the relative distances between classes cannot arise when there is only one such distance. With three-class solutions, however, the problem of relative spacing is a critical one. Without some resolution of it the regressions of tests on the latent continuum could not be drawn. Nor could the shape of the distribution of positions along the latent continuum be ascertained. In both of the present three-class solutions, the problem was resolved by the *arbitrary* assumption of equal spacing of the classes along the latent continuum. The regressions were drawn on that basis, and on the same basis the latent distribution in each case became symmetric and approximately normal. Other assumptions about the underlying metric would have led to different regressions and to different latent distributions. A separate paper [9], stemming from some recent developments in latent structure analysis [16], deals further with this metric problem. It indicates one way in which, with the aid of manifest product moments of still higher order, a metric for the latent continuum can be made to emerge as an integral part of the latent profile solution.

A second limitation of all latent profile examples in the present paper is their unidimensionality. It will be recalled that nothing in the development of the latent profile equations restricts the number of underlying dimensions within which the latent classes lie. Of course the two-class examples here can be understood in terms of a single continuum, for that would be true of any two-class case. The present three-class examples, however, are unidimensional because of the special nature (homogeneous in content but graded in difficulty) of the tests involved in them. Many three-class examples would require two underlying dimensions for an adequate understanding of their psychological meaning. In general, a q-class solution could require as many as $(q - 1)$ underlying dimensions for its interpretation. Subsequent work [cf. 10] will deal with such multidimensional examples and with the problems of dimensionality and metric that arise in their interpretation.

The reader may have noticed that in all four latent profile examples no mention was made of a need for manifest data of order higher than the third. Even for the two empirical examples a unique solution would not have required the use of such higher order manifest data. These higher order data therefore constitute a means for testing the assumption of higher order within-class independence. This could be done by comparing the given higher order data with the corresponding fitted values as generated from the latent parameters by subsequent lines of equations (13) or (14). Alternatively it could be argued that, if the solution never requires data above third order, there is no need to postulate within-class independence beyond

that order. From this parsimonious viewpoint, the latent profile equations need extend no higher than third order, and it would be inappropriate to think of using higher order data to test the fit of the model. However, it might then be more important to test the adequacy of the solution by relating it to variables not included in the original analysis. (An empirical example in [8] gives an illustration of how this could be carried out in latent structure analysis, where the very same argument over the use or non-use of higher order equations not needed in a particular solution can be made.)

The latent profile equations that have been derived and illustrated here are analogous to only one form of latent structure analysis—that known as the discrete class case. The analysis divides the sample into a small number of discrete classes possessing second- and third-order within-class independence, and stops there. Other varieties of latent structure analysis, one of which has already been referred to, have gone further in stipulating the algebraic form of the regressions (the so-called *trace lines* [cf. 15] of latent structure analysis) or of the set of class sizes, or of both. Always, however, the postulate of within-class independence is retained. Usually these further restraints require within-class independence of higher than third order, so that the corresponding higher orders of manifest data become directly involved in the solution. Most of these variants of latent structure analysis are readily translated into latent profile terms. In fact, the analogy between the two models is so close that almost whatever progress is made in the various solutions for one model is convertible into a corresponding advance for the other.

Conclusion

After outlining the derivation of the factor analysis and latent structure models, this paper has shown how the latter can be generalized for analyzing the interrelations among quantitative measures in a way that avoids some of the troublesome problems of factor analysis. The resulting latent profile model is applied to some simple fictitious and empirical data to illustrate its use. Because such applications may seem to show some promise, it is perhaps appropriate and not premature to conclude this paper merely by broadening the reference for the admonition ([18], p. 70) with which it began.

> It would be unfortunate if some initial success with the analytical methods . . . described here should lead us to commit ourselves to them with such force of habit as to disregard the development of entirely different constructs that may be indicated by improvements in measurement and by inconsistencies between theory and experiment.

REFERENCES

[1] Anderson, T. W. On estimation of parameters in latent structure analysis. *Psychometrika*, 1954, 19, 1-10.

[2] Carroll, J. B. The effect of difficulty and chance success on correlations between items or between tests. *Psychometrika*, 1945, **10**, 1-19.
[3] Davis, F. B. Fundamental factors of comprehension in reading. *Psychometrika*, 1944, **9**, 185-197.
[4] Ferguson, G. A. The factorial interpretation of test difficulty. *Psychometrika*, 1941, **6**, 323-329.
[5] Gibson, W. A. Applications of the mathematics of multiple-factor analysis to problems of latent structure analysis. In P. F. Lazarsfeld et al., The use of mathematical models in the measurement of attitudes. Res. Memo. No. 455, Rand Corp., 1951.
[6] Gibson, W. A. Applications of the mathematics of multiple-factor analysis to problems of latent structure analysis. Unpublished doctoral dissertation, Univ. Chicago, 1951.
[7] Gibson, W. A. An extension of Anderson's solution for the latent structure equations. *Psychometrika*, 1955, **20**, 69-73.
[8] Gibson, W. A. Multiple factors and latent structure. Unpublished manuscript.
[9] Gibson, W. A. Nonlinear factor analysis, single factor case. *Amer. Psychologist*, 1955, **10**, 438. (Abstract)
[10] Gibson, W. A. Nonlinear factors in two dimensions. *Amer. Psychologist*, 1956, **11**, 415 (Abstract)
[11] Green, B. F. A general solution for the latent class model of latent structure analysis. *Psychometrika*, 1951, **16**, 151-166.
[12] Green, B. F. Latent structure analysis and its relation to factor analysis. *J. Amer. statist. Ass.*, 1952, **47**, 71-76.
[13] Guilford, J. P. The difficulty of a test and its factor composition. *Psychometrika*, 1941, **6**, 67-77.
[14] Lazarsfeld, P. F. The interpretation and computation of some latent structures. In S. A. Stouffer et al., *Measurement and prediction*. Princeton: Princeton Univ. Press, 1950. Ch. 11.
[15] Lazarsfeld, P. F. The logical and mathematical foundation of latent structure analysis. In S. A. Stouffer et al., *Measurement and prediction*. Princeton: Princeton Univ. Press, 1950. Ch. 10.
[16] Lazarsfeld, P. F. Some new results and problems in latent structure analysis. Unpublished manuscript.
[17] Lazarsfeld, P. F., et al. The use of mathematical models in the measurement of attitudes. Res. Memo. No. 455, Rand Corp., 1951.
[18] Thurstone, L. L. *Multiple-factor analysis*. Chicago: Univ. Chicago Press, 1947.
[19] Thurstone, L. L. Note on a reanalysis of Davis' reading tests. *Psychometrika*, 1946, **11**, 185-188.
[20] Wherry, R. J. and Gaylord, R. H. Factor pattern of test items and tests as a function of the correlation coefficient: content, difficulty, and constant error factors. *Psychometrika*, 1944, **9**, 237-244.

Manuscript received 2/7/58

Revised manuscript received 9/19/58

PSYCHOMETRIKA—VOL. 24, NO. 3
SEPTEMBER, 1959

STRATEGIES AND LEARNING MODELS

STEVEN J. BRYANT AND JOHN G. MARICA

FRESNO STATE COLLEGE

A class of strategies is defined, each member of which possesses a certain plausibility. If a subject follows any strategy in this class in a two-choice learning experiment of the type dealt with by the Estes model, the subject's long-run behavior will be the same as that predicted by the Estes model.

Consider a partial reinforcement experiment in which there are two alternatives of behavior A_1 and A_2 . If A_1 is chosen on a particular trial, reward occurs with probability π_1 ; if A_2 is chosen, reward occurs with probability π_2 . Let $p_1(t)$ be the probability the subject chooses A_1 on trial t.

In special cases of the Estes stimulus-sampling model Estes ([1], ch. 9, p. 134) has found that if $\bar{p}_1(t)$ is the expected mean probability of choosing A_1 then

$$(1) \quad \bar{p}_1(n+1) = \theta(1-\pi_2) + (1 - 2\theta + \theta\pi_1 + \theta\pi_2)\bar{p}_1(n), \quad (\theta \text{ a constant})$$

and, letting $\bar{p}_1 = \lim_{t\to\infty} \bar{p}_1(t)$ then,

$$(2) \qquad \bar{p}_1 = \frac{1-\pi_2}{2-\pi_1-\pi_2}.$$

Since the subject can "do better" (in terms of maximizing his expectation) by always choosing the alternative having greater reward probability, questions arise as to the rationality of this type of behavior (Flood [1], ch. 18).

Simon [2] has shown that the Estes result can be derived from the assumption that the subject is behaving rationally in a certain game-theoretic sense and is attempting to minimize his "regret."

In this note,

$$\lim_{t\to\infty} p_1(t) = \frac{1-\pi_2}{2-\pi_1-\pi_2}$$

is derived from the assumption that the subject is adhering to one of a certain class of strategies $\{S_\tau\}$. A member of this class may be described by having the subject decide in advance on a policy, i.e.:

(i) when a choice is followed by reward, the same choice will be made on the next trial *all the time* (with probability 1);

(ii) when a choice is not followed by reward, the same choice will be made on the next trial with some fixed probability $\gamma \neq 1$.

The above axioms may be interpreted in several ways, e.g., the subject equips himself with a mechanism such as a barrel of balls, γ of which are marked "repeat," and he samples the barrel with replacement each time he is not rewarded, to decide what to do on the next trial. Or the subject has some built-in mechanism which affects his behavior in the same way. At any rate, each real number γ, $0 \leq \gamma < 1$, determines a strategy, call it S_γ.

The equation

$$p_1(n) = \pi_1 p_1(n-1) + (1-\pi_1)p_1(n-1)\gamma + (1-\pi_2)[1-p_1(n-1)](1-\gamma) \tag{3}$$

may be regarded as a formal statement of axioms (i) and (ii), for if the subject is following strategy S_γ he will choose A_1 on trial n:

(a) if he chose A_1 on the trial $n-1$ and was rewarded, (which occurs with probability $\pi_1 p_1 (n-1)$);

(b) if he chose A_1 on the trial $n-1$ and was not rewarded, and the mechanism tells him to choose A_1 on the next trial, (which occurs with probability $(1-\pi_1)p_1(n-1)\gamma$);

(c) if he chose A_2 on the trial $n-1$, was not rewarded, and the mechanism tells him to choose A_1 on the next trial, (which occurs with probability $(1-\pi_2)\,[1-p_1(n-1)]\,(1-\gamma)$).

The above difference equation is of the same type as the one obtained from the stimulus-sampling model and

$$\lim_{t\to\infty} p_1(t) = \frac{1-\pi_2}{2-\pi_1-\pi_2}. \tag{4}$$

This time, the limit is independent of γ and so each member S_γ of the class $\{S_\gamma\}$ yields the same limiting behavior in agreement with the Estes model as well as that of Bush, Mosteller, and Thompson ([1], ch. 8).

Except for the bars over $p_1(n)$ and $p_1(n-1)$, (1) is the same as (3) if and only if $\theta = 1-\gamma$. From this, one may conclude that, if a subject follows one of the strategies S_γ defined by (1) and (2), his trial-by-trial behavior will be the same as the mean behavior predicted by the stimulus-sampling model if and only if γ, the probability of repeating after nonreward, is equal to $1-\theta$; to stretch a point, the constant value attributed to θ may be paraphrased "hope springs eternal, with probability $1-\theta$."

A Generalization

Each member of the class of strategies just indicated determines the subject's choice on the basis of what happened on the preceding trial. There

are also strategies in which the subject takes into account what has happened on the preceding k trials, which also lead to the same asymptote.

To construct such strategies, let k be a fixed positive integer. Also let $w_1, \cdots, w_k$ be k nonnegative real numbers such that $\sum_{i=1}^{k} w_i = 1$. Let $\gamma_1, \cdots, \gamma_k$ be k real numbers such that $0 \leq \gamma_i < 1$. Let

$$\delta_i = \pi_1 p_1(n-i) + (1-\pi_1)p_1(n-i)\gamma_i + (1-\pi_2)[1-p_1(n-i)](1-\gamma_i).$$

Then let

$$p_1(n) = \sum_{i=1}^{k} w_i \, \delta_i .$$

This not only defines $p_1(n)$, the probability of choosing A_1 on the trial n, but yields a strategy, i.e., the subject looks at the last k trials, each of which has a "weight." Thus, if the subject chose A_1 on the trial $n-i$ and was rewarded, that trial contributes weight w_i. If the subject chose A_1 and was not rewarded on the trial $n-i$, then that trial contributes weight $w_i\gamma_i$. If the subject chose A_2 and was not rewarded on the trial $n-i$, that trial contributes weight $w_i(1-\gamma_i)$. Then the probability of choosing A_1 on the trial n, given a particular sequence of k preceding outcomes, is the sum of the weights contributed. For example, if $\gamma_1 = \gamma_2 = \cdots = \gamma_k = 0$, then each trial contributes weight w_i if the subject chose A_1 and was rewarded, or chose A_2 and was not rewarded; otherwise, the trial contributes 0 weight to the probability of choosing A_1 on the trial n.

Now, it is easy to show that

$$\lim_{n\to\infty} p_1(n) = \frac{1-\pi_2}{2-\pi_1-\pi_2}.$$

Proof. $\lim_{n\to\infty} p_1(n)$ exists. Let $\lim_{n\to\infty} p_1(n) = p_1$; then p_1 is the solution of the equation

$$p_1 = \pi_1 p_1 \sum_{i=1}^{k} w_i + (1-\pi_1)p_1 \sum_{i=1}^{k} w_i\gamma_i + (1-\pi_2)(1-p_1)\sum_{i=1}^{k} w_i(1-\gamma_i).$$

Let

$$\gamma = \sum_{i=1}^{k} w_i\gamma_i ;$$

then

$$p_1 = \pi_1 p_1 + (1-\pi_1)p_1\gamma + (1-\pi_2)(1-p_1)(1-\gamma)$$

and $\gamma \neq 1$, and hence

$$p_1 = \frac{1-\pi_2}{2-\pi_1-\pi_2}.$$

This shows that, *if the subject follows a strategy S_γ in the class $\{S_\gamma\}$*, his asymptote will be the same regardless of how good his "memory" is. In short, he will do no better in the long run by considering, say; the last 100 trials each time, than if he only considers the last trial; and furthermore, his long-run behavior is unaffected by the choice of weights which he gives to each trial.

REFERENCES

[1] Thrall, R. M., Coombs, C. H., and Davis, R. L. (Eds.) *Decision processes.* New York: Wiley, 1954.

[2] Simon, H. A. A comparison of game theory and learning theory. *Psychometrika,* 1956, 21, 267-272.

Manuscript received 10/6/58

Revised manuscript received 1/12/59

PSYCHOMETRIKA—VOL. 24, NO. 3
SEPTEMBER, 1959

A NOTE ON TRYON'S MEASURE OF RELIABILITY

ALLEN L. EDWARDS

THE UNIVERSITY OF WASHINGTON

Two alternative formulas, based upon the analysis of variance, are given for Tryon's general form for the reliability coefficient.

A general form for the reliability coefficient r_{tt} of an unstratified composite X_t, consisting of the sum of n observations obtained for each of N individuals, has been published by Tryon [2]. For example, for an $N \times n$ matrix of observations, as shown in Table 1, with N individuals or rows and n columns of observations, the general form is given by

$$r_{tt} = \frac{n^2 \bar{c}_{ij}}{V_t}, \tag{1}$$

where $\bar{c}_{ij}$ is the average of the covariances of the entries in the n columns, and V_t is the variance of the N row sums.

Tryon shows that r_{tt} may be calculated from any one of a number of algebraically identical formulas, but does not show the relationship between (1) and two useful alternatives based upon the analysis of variance. For example, (1) can also be written, in a form given by Hoyt [1], as

$$r_{tt} = 1 - \frac{\mathrm{MS}_{re}}{\mathrm{MS}_r}, \tag{2}$$

where MS_{re} is the row $\times$ column interaction mean square and MS_r is the mean square between rows, or as

$$r_{tt} = \frac{n}{n-1}\left(1 - \frac{\mathrm{MS}_{we}}{\mathrm{MS}_r}\right), \tag{3}$$

where MS_{we} is the mean square within columns.

For the data of Table 1 Tryon reports r_{tt}, based upon (1), as .880. In Table 2 appears the analysis of variance for the same example. Substituting in (2) with the appropriate values from Table 2 yields

$$r_{tt} = 1 - \frac{3.728}{31.122} = .880.$$

For the same example, the sum of squares within columns is 280.1 +

TABLE 1*

Illustrative Score Matrix

Subjects	Test Samples					
	X_1	X_2	X_3	X_4	X_5	X_t
1	6	2	1	0	0	9
2	8	6	5	2	4	25
3	10	12	7	7	7	43
4	5	11	11	9	8	44
5	6	3	0	0	1	10
6	11	7	9	6	1	34
7	7	7	2	5	5	26
8	4	7	4	4	1	20
9	6	3	3	2	4	18
10	6	5	1	3	1	16
Σx_i	69	63	43	38	32	245

*From Tryon [2]

TABLE 2

Analysis of Variance of the Scores Given in Table 1

Source of variation	Sum of squares	df	Mean square
Between rows (subjects)	280.1	9	31.122
Between columns (scores)	104.2	4	26.050
Row x column interaction	134.2	36	3.728
Total	518.5	49	

134.2 = 414.3, with $n(N-1)$ degrees of freedom. MS_{wc} is, therefore, equal to 414.3/45 = 9.207. Substituting with this value in (3),

$$r_{tt} = \frac{5}{5-1}\left(1 - \frac{9.207}{31.122}\right) = .880.$$

The two analysis of variance formulas will be shown to be algebraically equivalent to (1). Let the sum of squares within columns be $\sum x_{wc}^2$, the sum of squares between rows be $\sum x_r^2$, and the row × column sum of squares be $\sum x_{rc}^2$, with degrees of freedom of $n(N-1)$, $(N-1)$, and $(n-1)(N-1)$, respectively. Division of each sum of squares by its degrees of freedom gives the corresponding mean squares MS_{wc}, MS_r, and MS_{rc}.

The mean of the variances within columns is

$$\bar{V}_i = (s_1^2 + s_2^2 + \cdots + s_n^2)/n = \frac{\sum x_{wc}^2}{n(N-1)} = MS_{wc}. \tag{4}$$

Tryon's V_t is the variance of the row *sums* and, in his notation, is

$$V_t = n\bar{V}_i + n(n-1)\bar{c}_{ij}. \tag{5}$$

The variance of the row *means* will then be

$$s_{\bar{x}}^2 = \frac{1}{n^2}[n\bar{V}_i + n(n-1)\bar{c}_{ij}]. \tag{6}$$

Also, $ns_{\bar{x}}^2 = MS_r$ and, therefore,

$$n^2 s_{\bar{x}}^2 = n\,MS_r = V_t. \tag{7}$$

Solving for $\bar{c}_{ij}$ in (5) and substituting with this value in (1), gives Tryon's variance form or

$$r_{tt} = \frac{n}{n-1}\left(1 - \frac{\sum V_i}{V_t}\right). \tag{8}$$

To derive formula (3), substitute from (4) and (7) in (8) to obtain

$$r_{tt} = \frac{n}{n-1}\left(1 - \frac{MS_{wc}}{MS_r}\right). \tag{9}$$

Formula (2) may be derived by noting that $\sum x_{wc}^2 = \sum x_{rc}^2 + \sum x_r^2$. Then, substituting in (9)

$$r_{tt} = \frac{n}{n-1} - \frac{(\sum x_{rc}^2 + \sum x_r^2)/(N-1)}{(n-1)\,MS_r}$$

$$= 1 - \frac{MS_{rc}}{MS_r}.$$

REFERENCES

[1] Hoyt, C. Test reliability obtained by analysis of variance. *Psychometrika*, 1941, **6**, 153-160.

[2] Tryon, R. C. Reliability and behavior domain validity: reformulation and historical critique. *Psychol. Bull.*, 1957, **54**, 229-249.

Manuscript received 8/16/58

Revised manuscript received 2/9/59

PSYCHOMETRIKA—VOL. 24, NO. 3
SEPTEMBER, 1959

AN IMPROVED PROCEDURE FOR THE WHERRY-WINER METHOD FOR FACTORING LARGE NUMBERS OF ITEMS

LEROY WOLINS

IOWA STATE COLLEGE

A technique is presented that differs from the previous one in that the use of variance terms is eliminated from the computations; thus some formulas are simplified. A rationale for the improved method is presented.

In a previous article [1] the following formula is recommended for successive use to secure estimates of factor loadings:

$$r_{iK} = \frac{\left(r_{iK'}\frac{\sigma_{K'}}{\bar{\sigma}_i} - 1\right) + h_{ii}^2}{\sqrt{\left(\frac{\sigma_{K'}^2}{\bar{\sigma}_i^2} - n_{K'}\right) + \sum h_{ii}^2}}. \tag{1}$$

In (1) r_{iK} is the factor loading of an item. The difference between $r_{iK'}$ and r_{iK} is that the latter is based on computations involving communalities whereas the former is based on the same correlation matrix but with unity entered in each diagonal cell. The standard deviation of a total score based on all of the items included in a cluster is $\sigma_{K'}$, and the average standard deviation of the items included in cluster K is $\bar{\sigma}_i$. The number of items included in cluster K is $n_{K'}$ and, of course, represents the number of ones in the diagonal of the matrix of item intercorrelations.

By making substitutions in (1) suggested by the relationships

$$\frac{\sigma_{K'}}{\bar{\sigma}_i} = \sqrt{\sum\sum r_{ij}}, \tag{2}$$

and

$$\sum r_{iK'} = \sum\left[\frac{\sum r_{ij}}{\sqrt{\sum\sum r_{ij}}}\right] = \sqrt{\sum\sum r_{ij}}, \tag{3}$$

(1) becomes

$$r_{iK} = \frac{(r_{iK'}\sum r_{iK'} - 1) + h_{ii}^2}{\sqrt{(\sum r_{iK'})^2 - n_{K'} + \sum h_{ii}^2}}. \tag{4}$$

As in the previous method the correlations of items not in the cluster are converted to factor loadings by multiplying them by $C_{K'}$ and

$$r_{KL} = C_{K'}C_{L'}r_{K'L'} , \tag{5}$$

but by the present method

$$C_{K'} = \frac{\sum r_{iK'}}{\sum r_{iK}}. \tag{6}$$

The rationale for this modification is discussed in connection with the following identity:

$$r_{aK'} = \frac{[r_{\sigma_a r_{aj}}][\sigma_{\sigma_i}][\sigma_{r_{aj}}] + \bar{\sigma}_i \bar{r}_{aj}}{\sqrt{[r_{(\sigma_i \sigma_j) r_{ij}}][\sigma_{\sigma_i \sigma_j}][\sigma_{r_{ij}}] + \bar{\sigma}_i^2 \bar{r}_{ij}}} , \tag{7}$$

where

$r_{aK'}$ = the correlation of a particular item, a, with the cluster that contains it,

$r_{\sigma_a r_{aj}}$ = the correlation of a particular column or row of item intercorrelations with the column or row of item standard deviations,

σ_{σ_i} = the standard deviations of the item standard deviations,

$\sigma_{r_{aj}}$ = the standard deviation of a particular row or column of item intercorrelations,

$\bar{\sigma}_i$ = the mean standard deviation of the items,

$\bar{r}_{aj}$ = the mean of a particular column or row of item intercorrelations,

$r_{(\sigma_i \sigma_j) r_{ij}}$ = the correlation of the product of all combinations and permutations of two item standard deviations with their respective item intercorrelations,

$\sigma_{\sigma_i \sigma_j}$ = the standard deviation of the product of all combinations and permutations of two item standard deviations,

$\sigma_{r_{ij}}$ = the standard deviation of all item intercorrelations,

$\bar{r}_{ij}$ = the mean of all item intercorrelations.

The pertinent aspect of (7) is that it illustrates how it might be reduced to the simplified form suggested in the present paper. Such simplification is realized when any of the three bracketed terms in both the numerator and denominator is zero or is near zero.

The first terms to consider are $\sigma_{r_{aj}}$ and $\sigma_{r_{ij}}$. With respect to manipulating the data so as to make these terms zero, nothing can be done. It is expected that the correlations within a correlation matrix or within a row or column of such a matrix will vary, but these standard deviations will certainly be less than one, and if the clustering was done well, these standard deviations will be quite small.

The σ_{σ_i} and $\sigma_{\sigma_i \sigma_j}$ terms can be manipulated. It is conceivable with the

use of electronic computors that the distribution of item responses to each item could be economically transformed to standard scores prior to computing the sums to obtain cluster scores. Later comments will point out that this is probably not justified.

The $r_{\sigma_i r_{\sigma j}}$ and $r_{(\sigma_i \sigma_j) r_{ij}}$ terms will be zero on the average, theoretically, if tetrachorics are used and, theoretically, will be positive on the average if phi coefficients are used, varying in size depending on σ_{σ_i} and $\sigma_{\sigma_i \sigma_j}$. However, even if chance were the only thing contributing to the variability of $r_{\sigma_i r_{\sigma j}}$ and $r_{(\sigma_i \sigma_j) r_{ij}}$, the user of the present technique would not wish to depend on any theoretical average value since the standard error of these coefficients would be quite large with the small number of observations which would occur in those problems requiring iteration.

In the Wherry-Winer paper the assumption made was that σ_{σ_i} is zero. It was also assumed, tacitly, that the bracketed terms in the denominator of (7) are the same for tetrachorics as for phis. Since both $r_{\sigma r}$ and σ_r will change as a result of using tetrachorics, this latter tacit assumption is not justified; however, since all three terms in both the denominator and the numerator will be less than one and probably closer to zero than to one, the product of the three bracketed terms will probably be very small. This appears to be the reason Wherry and Winer found close correspondence of their method to results obtained through actually extracting a centroid. This becomes evident when one considers that if items included in a cluster vary in difficulty as much as from .2 to .8, σ_{σ_i} will be in the order of .02 and $\sigma_{\sigma_i \sigma_j}$ will be even smaller.

Thus, it is concluded on a rational basis that the present computational procedure is an improvement of the former one, certainly with respect to computational ease and certainly when tetrachorics are used. With respect to phis, it is difficult to determine if the present method is better or worse than the former one. One can say with certainty that there will not be much difference between them on the average since the product of the bracketed terms will be small in most cases.

A third method available, discussed in [2], is clearly superior to both methods discussed above, in that the results of this third method are identical to those obtained from extracting a centroid from a cluster. An objection to the general use of this exact method is that it requires more work. The exact method should be used on those occasions when iteration leads to divergence and the investigator wishes to salvage the cluster. There is no guarantee that the exact method will salvage the cluster, of course, since divergence can be caused by a negative r_{ij} as well as by large values of the bracketed terms.

Another use of the exact method is when the items are not dichotomous. If items are responded to on a 1 to 5 scale, say, it may be easier to use the exact method than to dichotomize.

The computational formula for the exact method is

$$\hat{r}_{jK} = \frac{\sigma_j(r_{jK'} \sum \sigma_i r_{iK'} - \sigma_j) + \hat{h}_{jj}^2}{\sqrt{\sum [\sigma_j(r_{jK'} \sum \sigma_i r_{iK'} - \sigma_j) + \hat{h}_{jj}^2]}}, \tag{8}$$

where

$$r_{jK} = \frac{\hat{r}_{jK}}{\sigma_j}. \tag{9}$$

REFERENCES

[1] Wherry, R. J. and Winer, B. J. A method for factoring large numbers of items. *Psychometrika*, 1953, **18**, 161-179.

[2] Winer, B. J. Iterative factor analysis: its psychological and mathematical bases. Unpublished doctoral dissertation, Ohio State Univ., 1952.

Manuscript received 3/27/58

Revised manuscript received 3/11/59

PSYCHOMETRIKA—VOL. 24, NO. 3
SEPTEMBER, 1959

GENERATING VARIABLES WITH ARBITRARY PROPERTIES

PAUL J. HOFFMAN

UNIVERSITY OF OREGON

There are occasions in psychological research where it is desirable to have available sets of variables with arbitrary intercorrelations. A quite simple procedure is described for generating pairs of such variables.

There are a number of instances in research and teaching where it is desired to produce fictitious data that exhibit particular characteristics. Linear transformation of scores on an existing variable can easily produce a variable with any desired mean and standard deviation. In test construction, where item information is available, the items can be so selected as to yield a test of given difficulty and reliability [1]. It also has been shown [2] that a test can be constructed to yield a given weight in relation to a second test or a composite.

This note describes a procedure whereby pairs of variables may be constructed from a table of random numbers so that their correlation will be of any given predetermined magnitude. The operations can be easily carried out in a few minutes on a desk calculator.

Let X and Z be random variables with the restriction $r_{xz} = 0$. We wish to determine a distribution Y such that r_{xy} is of some arbitrary size. Let $Y_i = X_i + bZ_i$. In deviation score notation,

$$r_{xy} = \sum_{i=1}^{N} \frac{(x_i)(x_i + bz_i)}{N\sigma_x\sigma_{(x+bz)}} = \frac{\sigma_x^2 + br_{xz}\sigma_x\sigma_z}{\sigma_x\sigma_{(x+bz)}}$$

which, with $r_{xz} = 0$ becomes

$$r_{xy} = \frac{\sigma_x}{\sigma_{(x+bz)}}. \tag{1}$$

But

$$\sigma_{x+bz}^2 = \sigma_x^2 + b^2\sigma_z^2 . \tag{2}$$

Therefore, by substitution,

$$r_{xy}^2 = \frac{\sigma_x^2}{\sigma_x^2 + b^2\sigma_z^2},$$

or, solving for b,

$$b^2 = \frac{\sigma_x^2(1 - r_{xy}^2)}{\sigma_z^2 r_{xy}^2}, \qquad b = \frac{\sigma_x}{\sigma_z}\sqrt{\frac{1 - r_{xy}^2}{r_{xy}^2}};$$

since $\sqrt{1-r_{xy}^2} = k_{xy}$ = the coefficient of alienation,

$$b = \frac{k_{xy}\sigma_x}{r_{xy}\sigma_z}, \tag{3}$$

from which the Y_i can be readily computed.

The Y distribution which results will have a mean,

$$\bar{Y} = \bar{X} + b\bar{Z}$$

TABLE 1

Two Normal Variables, X and Z. $\bar{X} = \bar{Z} = 5.00$, $\sigma_X = \sigma_Z = 1.00$, and $r_{XZ} = .00$

X	Z	X	Z	X	Z	X	Z
5.966	4.249	5.424	5.404	5.188	5.303	5.044	3.711
5.613	5.741	6.236	4.793	8.044	4.852	4.357	5.166
5.339	4.452	3.979	6.024	4.940	6.163	4.989	4.487
2.539	3.733	4.388	5.726	4.709	5.449	6.942	4.794
4.414	5.851	5.160	6.472	3.712	5.460	4.351	4.921
5.727	3.781	5.499	5.112	3.598	3.677	2.854	5.176
5.334	6.403	6.270	5.068	3.595	5.494	5.402	6.906
4.615	7.093	6.166	5.053	5.821	5.635	4.425	5.104
5.953	4.690	5.078	6.324	3.529	4.556	4.325	5.603
6.527	6.158	5.533	2.699	4.269	5.120	3.818	6.525
5.061	3.571	4.174	5.414	5.691	3.583	5.808	3.862
5.280	4.093	2.927	6.493	4.966	4.904	6.174	3.949
5.041	5.088	5.003	4.229	4.993	4.304	5.722	6.794
2.973	4.690	5.792	5.720	4.893	6.666	3.816	5.129
5.169	4.966	6.186	5.442	3.880	5.436	4.129	5.458
3.714	4.158	4.364	6.345	5.511	5.737	5.097	4.593
5.929	3.976	3.789	5.193	4.853	5.269	3.866	3.202
6.138	6.024	3.671	4.795	4.473	3.288	5.196	6.876
5.832	5.868	5.704	3.971	6.256	5.371	4.191	5.037
6.660	4.126	4.365	5.208	4.891	3.955	5.503	3.845
5.510	3.924	4.768	6.919	3.756	4.723	6.346	5.028
5.030	3.810	3.949	3.760	5.069	3.927	4.278	5.138
5.983	3.050	5.738	5.736	3.728	4.463	4.669	4.980
4.849	5.119	7.519	5.064	5.575	4.773	6.041	5.607
5.586	6.887	4.657	4.397	4.156	2.418	5.440	4.722

and a standard deviation,

$$\sigma_y = \sqrt{\sigma_x^2 + b^2\sigma_z^2}\ .$$

If it is desired that the Y distribution have some arbitrary mean, $\bar{Y}_A$, and standard deviation, σ_{y_A}, as well as an arbitrary correlation with X, the individual scores may be computed from the formula

$$Y_i = \frac{\sigma_{y_A}}{\sigma_y}(X_i + bZ_i) + C$$

where σ_{y_A} = the desired standard deviation,
σ_y = the obtained standard deviation, $\sqrt{\sigma_x^2+b^2\sigma_z^2}$

$$C = \bar{Y} - \frac{\sigma_{y_A}}{\sigma_y}(\bar{X} + b\bar{Z})$$

and b is defined as in (3) above.

In the special case, where

$$\bar{X} = \bar{Z}, \qquad \sigma_x = \sigma_z = 1.00, \quad \text{and} \quad r_{xz} = 0,$$

we find by substitution that

$$b = k_{xy}/r_{xy}\ , \qquad \bar{Y} = \bar{X}(1 + [k_{xy}/r_{xy}]), \qquad \sigma_y = 1/r_{xy}\ .$$

Table 1 contains, for $N = 100$, a sample pair of normal variables (X and Z) such that $\bar{X} = \bar{Z} = 5.00$, $\sigma_x = \sigma_z = 1.00$, and $r_{xz} = .00$. Additional variables may be obtained from the author by request.

REFERENCES

[1] Gulliksen, H. *Theory of mental tests.* New York: Wiley, 1950.
[2] Hoffman, P. J. Predetermination of test weights. *Psychometrika*, 1958, **23**, 85-92.

Manuscript received 1/2/59

PSYCHOMETRIKA—VOL. 24, NO. 3
SEPTEMBER, 1959

A NOTE ON THE TRYON-KAISER SOLUTION FOR THE COMMUNALITIES

HENRY F. KAISER

UNIVERSITY OF ILLINOIS

The Tryon-Kaiser solution for the communalities is reviewed. Numerical investigation suggests that the procedure is applicable if and only if the correlation matrix has unique minimum rank communalities. This implies that this approach to the communality problem is not general enough to be of practical use.

In his recent review of the notion of communality from a cluster-analytic viewpoint, Tryon derives a formula "for the exact value of h^{2}" ([6], eq. 21). This treatment is interesting theoretically because it does not explicitly consider the dimensionality of the common-factor space. It is interesting practically because—at least for one example—it succeeded in obtaining an exact solution for the communalities.

Simultaneously and essentially independently, Kaiser [3] from a traditional factor-analytic point of view developed what ultimately is the same approach. His treatment consists of a derivation and an attempt to use the equations

$$\tilde{h}_j^2 = \hat{h}_j^2 - 1/\hat{r}^{jj} \qquad (j = 1, 2, \cdots, n),$$

to provide an iterative solution for the communalities, where $\hat{h}_j^2$ is a trial value for the jth communality, $\tilde{h}_j^2$ is a new (and hopefully improved) approximation to a solution, $\hat{r}^{jj}$ is the jth diagonal element of the inverse of the reduced correlation matrix with $\hat{h}_j^2$ in the diagonal, and n is the number of observed variables. This formula is applied as an attempt to compute the squared multiple correlation of test j on the remaining arbitrarily large number of tests in the hypothetical domain of content under consideration, a value which under very general conditions may be shown to equal the communality [2].

This note reports results obtained when this method of solving for the communalities was applied more extensively. In addition to Tryon's successful example, exact communalities for three further matrices were easily solved, using Kaiser's equations. These three examples, like Tryon's, had the property that the number of common factors was less than half the number of tests, as the off-diagonal elements of the correlation matrices involved had been generated artificially by multiplying an arbitrary factor matrix with this property by its transpose. For a second group of examples—six correlation

matrices based on empirical data—this method did not yield communalities; the iterative procedure failed to converge.

The obvious question is whether this method will succeed only for artificial correlation matrices and not for data from the real world. The answer probably lies in a theorem of Albert's and in some results of Ledermann. If r is the number of common factors, Albert [1] proved that when $r < n/2$, there exist *unique* communalities such that the resulting reduced correlation matrix has rank r. If it is postulated that the Tryon-Kaiser procedure will be applicable if the correlation matrix has unique minimum rank communalities, Albert's theorem would account for the success with artificial matrices. On the other hand, *empirical correlation matrices do not have unique minimum rank communalities.* This follows from Ledermann [4]. He has shown that if

$$r \geqq \tfrac{1}{2}(2n + 1 - \sqrt{8n + 1}),$$

the communalities will *not* be unique. He has also shown that if r is to be less than $\frac{1}{2}(2n + 1 - \sqrt{8n + 1})$, special conditions must hold exactly among the off-diagonal elements of the correlation matrix. Because sample correlation coefficients are continuous random variables, Ledermann's special conditions can only hold with probability zero, and consequently in practice, unique communalities may occur only with zero probability.

By systematically varying r/n in constructing additional artificial correlation matrices, extensive further numerical investigation uniformly confirmed the hypothesis that the Tryon-Kaiser solution for the communalities will converge if and only if the correlation matrix under consideration has unique minimum rank communalities. Indeed, unique negative "communalities" from non-Gramian matrices (generated with imaginary factors) were easily found. Attempts to prove the hypothesis algebraically with methods described by Scarborough ([5], pp. 209–211) have not been successful.

The difficulty with the Tryon-Kaiser solution is that it is incomplete. What is needed is a criterion for selecting among the inevitable multiple solutions for empirical correlation matrices. What this criterion might be seems a difficult scientific (not mathematical or statistical) problem. It does not appear to have been explored systematically.

REFERENCES

[1] Albert, A. A. The minimum rank of a correlation matrix. *Proc. nat. Acad. Sci.*, 1944, **30**, 144-148.

[2] Guttman, L. "Best possible" systematic estimates of communalities. *Psychometrika*, 1956, **21**, 273-285.

[3] Kaiser, H. F. Solution for the communalities: a preliminary report. Rep. No. 5, Contract No. AF 41(657)-76, Univ. California, 1956.

[4] Ledermann, W. On the rank of the reduced correlational matrix in multiple factor analysis. *Psychometrika*, 1937, **2**, 85-93.

[5] Scarborough, J. B. *Numerical mathematical analysis.* (2nd ed.) Baltimore: Johns Hopkins, 1950.
[6] Tryon, R. C. Communality of a variable: formulation by cluster analysis. *Psychometrika,* 1957, **22,** 241-260.

Manuscript received 1/7/58

Revised manuscript received 3/9/59

PSYCHOMETRIKA—VOL. 24, NO. 3
SEPTEMBER, 1959

A NOTE ON THE USE OF TRIADS FOR PAIRED COMPARISONS

R. E. SCHUCKER
PURDUE UNIVERSITY

When scaling a large number of stimuli from comparative judgments, considerable savings in time and labor may be realized if stimuli are presented in triad form rather than in pairs. If, for N stimuli, the proper configuration of triads can be assembled so that all possible pairs appear once, the paired judgment matrix may be reproduced with one-third fewer judgments and two-thirds fewer presentations than would be required with complete pairing. A simple procedure is described for enumerating triad configurations for which N is an odd multiple of three.

Application of the traditional method of paired comparisons rapidly becomes unwieldy as the number of stimuli increases beyond 20. However, the labor required in eliciting and analyzing large numbers of paired judgments has been greatly reduced with the advent of punched card procedures [3]. In addition, the burden on the subject has been eased through the development of a partial pairing technique [5, 6]. With partial pairing, the total number of pairs to be judged may be reduced any desired amount by pairing each stimulus with fewer than the $N - 1$ remaining stimuli.

Where the paired comparison method is applied for the purpose of developing an interval scale, the investigator may desire that all stimuli be scaled with equal accuracy. That is to say, it is required that each stimulus be compared *equally often* with all *remaining stimuli*. It is readily apparent that the partial pairing technique does not fulfill the requirements for balance with respect to the estimation of stimulus scale values. Hence, when large numbers of stimuli are to be scaled, an alternative to partial pairing is needed if scale values are to be estimated with equal accuracy and the volume of paired judgments kept within reasonable bounds.

One approach in reducing the total number of pairs in paired comparisons has been through the use of triads. The format of the Kuder Preference Record is a familiar example of the grouping of stimuli in three's rather than in pairs [4]. If for the triad A, B, C, judgments of the type "most" and "least" are obtained, with respect to the ordering of the objects along a psychological dimension, preferences for the pairs (A, B), (A, C) and (B, C) may be recovered. Similarly, if for N objects the proper configuration of triads be assembled, so that every possible pair appears exactly once, information for the complete, paired judgment matrix is obtained with only one-third as many triads and two-thirds as many judgments as would be required using pairs and traditional full pairing. For example, complete pairing for

an N of 57 requires 1596 pairs, while only $1596/3 = 532$ appropriate triads furnish the same paired information.

The problem remains of finding the triads needed for a given N in order that each element be paired once with every remaining element. The present paper touches upon a geometric solution that produces for certain N's a basic group of triads, from which the required configuration may be enumerated by cyclic permutation. Where the stimuli to be scaled are verbal in nature, the cycling procedure is readily adaptable to punched card equipment for the preparation of the triads and analysis of data.

Cyclic Enumeration of Configurations

In general, a configuration consists of N elements arranged in b sets of k elements each, with each element occurring in r sets, and each pair of elements occurring together in a set exactly λ times. The following relationships must hold.

$$\lambda = \frac{r(k-1)}{N-1}, \tag{1}$$

$$Nr = bk. \tag{2}$$

For the case under consideration $\lambda = 1$ and $k = 3$. Substituting these values in (1) and solving for r, yields $r = (N-1)/2$. Thus a configuration of triads satisfying the required relationships is possible only when N is odd. In addition, it should be pointed out that the conditions specified in (1) and (2) are necessary but not sufficient for the existence of a configuration. Having determined that a particular N satisfies the requirements, the task of assembling the triads of the configuration still remains.

Various methods for constructing configurations have been reviewed by Cox [2]. Class $\lambda = 1$ configurations have received considerable attention in the literature and numerous solutions are available for the subclass $k = 3$, where N is an odd multiple of three. Solutions for N of this form may be obtained by cycling basic groups of triads. Depending upon N, the enumeration of the complete configuration from the basic groups proceeds by 1-step cycles, 2-step cycles, 3-step cycles, or combinations of cycles. An example of enumeration by 1-step cycles is provided by the configuration for $N = 9$. From equations (1) and (2) it is seen that $r = 4$, $b = 12$. The configuration is completed by setting $N = 9$.

	Rep I	Rep II	Rep III	Rep IV
row	set	set	set	set
1	(1) 1 N 5	(4) 2 N 6	(7) 3 N 7	(10) 4 N 8
2	(2) 2 3 8	(5) 3 4 1	(8) 4 5 2	(11) 5 6 3
3	(3) 6 7 4	(6) 7 8 5	(9) 8 1 6	(12) 1 2 7

The triads are grouped into complete replications of the 9 elements. It will be observed that the remaining $r - 1$ triads in any row may be obtained

by a succession of one-step cyclic permutations on the elements of the triad appearing in the same row in Rep I. For example, the addition of 1 to each element in triad (3) gives triad (6). Similarly, by cycling the elements of triad (6), triad (9) is obtained, and so forth. Note that the second element of each of the row-1 triads is not cycled and that for all triads one re-cycles to 1 on the next step when an element equals $N - 1$. Since the complete configuration can be generated from Rep I, the sets of Rep I may be arbitrarily designated as solution triads.

A second configuration is shown to illustrate enumeration by two-step cycles. For this example, $N = 15$, $r = 7$, and $b = 35$.

Rep I				Rep II				Rep III				Rep IV			
set				set				set				set			
(1)	1	N	8	(6)	3	N	10	(11)	5	N	12	(16)	7	N	14
(2)	4	5	7	(7)	6	7	9	(12)	8	9	11	(17)	10	11	13
(3)	13	14	9	(8)	1	2	11	(13)	3	4	13	(18)	5	6	1
(4)	3	6	11	(9)	5	8	13	(14)	7	10	1	(19)	9	12	3
(5)	10	12	2	(10)	12	14	4	(15)	14	2	6	(20)	2	4	8

Rep V				Rep VI				Rep VII			
set				set				set			
(21)	9	N	2	(26)	11	N	4	(31)	13	N	6
(22)	12	13	1	(27)	14	1	3	(32)	2	3	5
(23)	7	8	3	(28)	9	10	5	(33)	11	12	7
(24)	11	14	5	(29)	13	2	7	(34)	1	4	9
(25)	4	6	10	(30)	6	8	12	(35)	8	10	14

Setting $N = 15$, the configuration satisfies equations (1) and (2). As before, the $r - 1$ remaining triads in any row can be generated starting with the corresponding triad of Rep I. For this example, however, cycling proceeds by 2-step increments. Thus, by adding 2 to each element of triad (2) one obtains triad (7), which gives rise in succession to triads (12), (17), (22), (27) and (32). Note that a re-cycling to 1 or 2 occurs on the next step for elements which equal $N - 2$ or $N - 1$.

Ball discusses variations on a geometric method for obtaining solution triads in configurations for which N is an odd multiple of three ([1], ch. 10). Essentially, the method involves the use of a circle with inscribed triangles, the points of which represent triad elements. The triangles of one replication are determined empirically, with the restriction that, when rotated within the circle, they generate the remaining replications of the configuration. Ball gives solution triads for $N = 9$, 15, 21, 27, 33, 39, 45, 51, 57, 63, 69, 75, 81, 87, 93 and 99.

Machine Cycling of Triads

The cycling procedures demonstrated in connection with the configurations for N's of 9 and 15 are readily adaptable to various card punch

calculators, which can be utilized for preparing the triads on IBM cards as well as for generating configurations. Judgments can be recorded directly on the cards and the analysis facilitated by using tabulating equipment. The program requirements for enumerating triad configurations are relatively straightforward and may be deduced directly from the two illustrative examples.* Either modular arithmetic or the conventional arithmetic shown may be employed, although use of the former results in a shorter, more general program capable of handling cyclic steps of any magnitude.

REFERENCES

[1] Ball, W. W. R. Revised by Coxeter, H. S. M. *Mathematical recreations and essays.* (11th ed.) London: Macmillan, 1939.

[2] Cox, G. M. Enumeration and construction of balanced incomplete block configurations. *Ann. math. Statist.*, 1940, **11**, 72–85.

[3] Kephart, N. C. and Oliver, J. E. A punched card procedure for use with the method of paired comparison. *J. appl. Psychol.*, 1952, **36**, 47–48.

[4] Kuder, F. G. *Examiner manual for the Kuder Preference Record.* Chicago: Sci. Res. Associates, 1953.

[5] McCormick, E. J. and Bachus, J. A. Paired comparison ratings. II: The reliability of ratings based on partial pairings. *J. appl. Psychol.*, 1952, **36**, 188–192.

[6] Oliver, J. E. A punched card procedure for use with partial pairing. *J. appl. Psychol.*, 1953, **37**, 129–130.

Manuscript received 8/1/58

Revised manuscript received 10/16/58

*A generalized cycling program for the IBM 604 Electronic Calculating Punch has been deposited as Document number 5840 with the ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. A copy may be secured by citing the Document number and by remitting $1.25 for photoprints, or $1.25 for 35 mm. microfilm. Advance payment is required. Make checks or money orders payable to: Chief, Photoduplication Service, Library of Congress.

BOOK REVIEWS

DONALD DAVIDSON, PATRICK SUPPES, AND SIDNEY SIEGEL. *Decision Making, An Experimental Approach.* Stanford: Stanford University Press, 1957. Pp. 121.

The basic problem which the authors have set out to tackle in the studies reported in this book is the separation of the effects of psychological probability and utility in decision making. They have considered this problem both in its theoretical framework and in the experimental verification of the theories set forth. The first chapter gives an introductory discussion of the problems of empirical interpretation of theories of decision making under uncertainty. The second chapter fills almost half of the book and deals with the basic model proposed by the authors. The third chapter reports an experiment which was designed to measure the cardinal utility of nonmonetary outcomes and to use the computed utilities to predict further choices. Two models were compared, one a linear programming model and the other an ordinal model based on straightforward comparisons. The linear programming model turned out to be considerably superior to the other and both were much superior to a random guessing method; moreover, if thresholds are ignored in order to obtain a larger number of predictions the accuracy remains significantly better than chance. The fourth chapter considers the problem of formulating utilities for incomparable outcomes; in contrast with the two preceding chapters the considerations here are entirely axiomatic in character.

The second chapter offers an explicit theory for the explanation of individual decision making under conditions of risk, and reports an experiment designed to test the theory in certain limited situations. The first step in the theory is to construct an event which has a psychological probability of one-half. Next, a set of six outcomes is constructed so as to be equally spaced in utility, and from these a utility function is constructed which is adequate to account for a certain class of preference and indifference relations. The experimental results lead the authors to conclude among other things that (1) the theory provides a practical approach to the problem of resolution of utility and psychological probability in situations involving risk; (2) under suitably controlled conditions certain people make choices among risky alternatives as though they were attempting to maximize expected utility, and (3) for such persons it is possible to construct a utility function unique up to a linear transformation. The point of departure for this model was the work of Mosteller and Nogee (An experimental measurement of utility, *Journal of Political Economy*, 1951, **59**, 371–404); connections with other previous work are traced and a short but well-chosen bibliography is included.

In my opinion this book will take a prominent place in the literature of decision making, but it is also clear that it does not represent the final word on any of the points considered. The expository level of the book is excellent; the prospective reader should be prepared for a reasonable amount of mathematical development of the axiom-definition-theorem type.

R. M. THRALL

University of Michigan

WARREN S. TORGERSON. *Theory and Methods of Scaling.* New York: John Wiley and Sons, 1958. Pp. xiii + 460.

In 1950 the Social Science Research Council appointed a Committee on Scaling Theory and Methods to review the status of scaling procedures in the social sciences. This committee came to the inevitable conclusion that a good survey of the recent and prolific work on scaling procedures was necessary. In 1951, Warren Torgerson, as a Research Associate of the Council, undertook the preparation of a monograph on scaling procedures.

After seven years, two of which constituted a long lost weekend in the Navy, the monograph had become a book and was published.

The result of Torgerson's and the Committee's efforts is a book which will be of considerable influence and of great value to social scientists. It is an excellent summary of the state of the art and theory of measurement in the social sciences, the sort of book of which there is a very real shortage in psychology and the social sciences in general. It was not written as an undergraduate textbook, and probably cannot be used as such. While written primarily as a reference book for technical workers, it does seem possible to use it at the graduate level if a reasonable mathematical background is assumed on the part of the student. I suspect it will be more widely used than was anticipated.

Whatever shortcomings the book has are due more to the state of the art than to incorrect handling of the issues by the author. Torgerson provides an excellent organization of the scaling methods as they are in fact used; the various methods are covered in sufficient detail to enable anybody to use a particular method after a thorough reading of the appropriate section of the book. While extensive mathematical treatment of the various methods is provided, the author does not remain solely at the abstract mathematical level, but introduces the reader to the realities of collecting data and treating them as required by a particular scaling technique. The book is, in other words, a happy combination of erudite sophistication and down-to-earth realism.

The book starts out with the usual introductory and organizing chapters—in this case three of them. The first two chapters cover the nature of measurement, types of measurement, etc. The third chapter organizes measurement as it occurs in psychology into three classes: (1) where subjects are scaled; (2) where the stimuli are scaled, and subject differences are attributed to sampling error; and (3) where both stimuli and subjects are scaled from the same set of data. The first type has not led to any important scaling developments and is largely ignored. The second type, called the *judgment approach*, and the third, called the *response approach*, form the basis for the organization of the rest of the book. The term *response* for the last approach is a little unfortunate, since it does not differentiate that approach from the second, in which responses are also made.

The next seven chapters deal with the judgment methods. Three chapters are devoted to subjective estimates, fractionation, and equisection methods—those methods which involve some appreciation on the part of the subject of numerical values on the subjective continuum under consideration. The next four chapters deal with the discriminative, or differential sensitivity methods—all those methods which are based primarily on the Thurstone models. The last three chapters are concerned with the response methods. One chapter is concerned primarily with the Guttman techniques, one is concerned with Lazarsfeld's latent-structure model, and the last is concerned with the techniques developed by Coombs and his students for dealing with comparative response data.

There is no need to go into the specific content of each chapter. Each deals with its subject matter in a detailed and thorough way. The only chapter which I wish had not been included is that which introduces the differential sensitivity methods, and treats briefly and lightly of the traditional psychophysical methods. In the rest of the book if Torgerson covered a subject at all, he covered it thoroughly. In this one chapter, however, the coverage of the psychophysical procedures is entirely inadequate, and it would have been more in keeping with the rest of the book not to discuss the subject at all.

My major over-all reactions to the book were concerned less with what Torgerson wrote than with the state of measurement theory and practice in psychology today. For example, Torgerson sets up quite clearly the different types of data matrix which are used in scaling work, and differentiates methods on the basis of the nature of the data. He makes clear the kinds of assumptions which are made with regard to both the stimulus and the subject variables in such matrices, and thus provides a more fundamental look

at the over-all picture than is customary. However, I found a desire to go back a step further, and to note that all basic sets of data involve three variables—stimulus, subject, and *response*—and that all of them have certain relations to the underlying true continuum. Just as each stimulus, or each subject, can be located on the continuum, so can each response—and there can be interactions between all three variables. It is not really necessary to assume anything fixed by the response, even a comparative response, and a truly general model for measurement would include solutions for the response as well as for stimuli and subjects. It is usual, of course, to have stimuli and subjects be orthogonal in experiments, while neither will normally be orthogonal to the response variable. It is quite possible, however, to give each subject a response and then tell him to find the stimulus which satisfies this response—just as is done with some fractionation procedures and with the equisection procedure. Any truly fundamental model for measurement must be able to deal with the relations between all three variables and the underlying continuum.

Actually, of course, we run into the reality of the lack of degrees of freedom for complete solutions, as Torgerson so often points out. And in fact the practical solutions with which we must deal often are such that an equivalent solution could have been obtained with simplifying assumptions other than those actually made. Using assumptions about the values or distributions of any of the three basic variables, we can solve for values of the other one or two. A change in the variables about which the assumptions are made will change the solutions as well. In other words, we cannot create more knowledge than the data give us; we can simply assign the knowledge to different variables.

In this frame of mind, I have one last reaction to report. It is that we have more sophistication about the nature of mathematical models of measurement than we do about experimental techniques for validating the models. Elegant scales can be constructed, but only after we have made enough assumptions to reduce the number of parameters to the number of available independent observations. Any verification of the assumptions requires the availability of more degrees of freedom, and experimental techniques must be devised which provide these degrees of freedom in a form appropriate to the assumptions being made. The mathematical models tell us only what *can* be so; better experimental techniques are necessary to tell us whether it *is* so.

These reactions are not intended as criticisms of the book, but rather as compliments to it. The book presents the whole range of material in sufficiently compact form that one is forced to try to get an overview. This fact, plus the over-all excellence of the presentation, will stimulate new and good research. I am tempted to say—and so I will say—that this book is a milestone.

WENDELL R. GARNER

The Johns Hopkins University

D. A. S. FRASER. *Nonparametric Methods in Statistics*. New York: John Wiley and Sons, 1957. Pp. x + 299.

Nonparametric Methods in Statistics (*NMS*) is an advanced work in statistical theory. *NMS* consists of two parts: an introduction to recent developments in the Neyman-Pearson tradition in statistical inference (Chapters 1 and 2) and an application of these developments to nonparametric statistics (Chapters 3, 4, 5, and 7). In addition, Chapter 6 is a survey of limit theorems useful in nonparametric theory.

The mathematical background necessary for comfortably reading *NMS* is a year's course in function theory. With less than advanced calculus the statements of many of the definitions and theorems are hardly intelligible. Mood (*Introduction to the Theory of Statistics*, McGraw-Hill, 1950) or preferably Cramér (*Mathematical Methods of Statistics*, Princeton University Press, 1946) are reasonable prerequisites in statistical theory. With-

out this background, *NMS* can possibly be used as a reference book for the theory. Siegel's book (*Nonparametric Statistics*, McGraw-Hill, 1956), at the other extreme, is nearly devoid of theory and mathematical content.

Besides the standard Neyman-Pearson optimum properties (e.g., most powerful, unbiased, and consistent), sufficiency, invariance, and completeness are stressed. These ideas are developed extensively and used in finding good nonparametric procedures—tests of hypotheses, point estimates, and tolerance intervals. Sufficiency and invariance have a strong intuitive appeal as criteria for optimality. Completeness is a mathematical condition that is useful when available. To illustrate these ideas consider the following problem. It is assumed that $X_1, \cdots, X_m, Y_1, \cdots, Y_n$ are mutually independent random variables. Assume all of the X's have a common distribution and all of the Y's have a common distribution (not necessarily the same as that of the X's). How should one estimate $\Pr(X_i < Y_j)$? ($\Pr(X_i < Y_j)$ appears in the study of the Wilcoxon two-sample procedure.)

It is clear that the (temporal) order in which the observations are made is irrelevant and attention can be restricted to $X_{(1)}, \cdots, X_{(m)}, Y_{(1)}, \cdots, Y_{(n)}$, where $X_{(1)}$ is the smallest of $X_1, \cdots, X_m$; $X_{(2)}$ is the second smallest, etc. In short, the order statistics form a sufficient statistic for the problem. The parameter of interest, $\Pr(X_i < Y_j)$, will have the same value whether the original random variables are used or whether any monotone increasing function (e.g., exponential of the random variable) is used. Therefore it is reasonable to restrict attention to those estimators that will not change when an arbitrary monotone transformation of the observations is applied. This is the principle of invariance. For this problem, invariance implies that the estimator must be a function of the ranks. Completeness implies that there is a unique unbiased estimator (and hence minimal variance unbiased estimator) which is a function of the invariant sufficient statistic (the ranks). It is the number of pairs (X_i, Y_j), where $X_i < Y_j$, divided by mn.

The approach of the above paragraph is applied to many estimating and testing experiments which arise in practice, e.g., (a) making inferences from a random sample about the location parameter of a distribution; (b) testing the null hypothesis that c samples come from the same distribution against the alternative that the c samples come from distributions differing in location only; (c) making inferences about the amount of dependence in a bivariate distribution function, i.e., testing for independence and estimating correlation; and (d) constructing tolerance sets from univariate and multivariate data.

The orientation towards Neyman-Pearson theory and linear models (analysis of variance, etc.) explains the lack of emphasis on tests of goodness of fit. In keeping with the theoretical orientation, many mathematical examples are given, and, on the other hand, applied examples and tables of distributions are not given.

I. Richard Savage

University of Minnesota

Philip J. McCarthy. *Introduction to Statistical Reasoning.* New York: McGraw-Hill, 1957. Pp. xiii + 402.

The author states his aims clearly in the Preface: ". . . a one-semester, nonmathematical course in statistics in which the instructor wishes to present a careful introduction to statistical reasoning. . . . A first course should emphasize the concepts of statistical reasoning rather than attempt to cover the wide variety of techniques. . . . Illustrative material should be drawn from investigations that are as significant as possible, and . . . has been chosen to range broadly over the social sciences. A very brief account of research problems usually accompanies each illustrative example and any student . . . may expect

to improve his insight into the problems of research methodology in the social sciences. . . . This selection of illustrative material from the social sciences has also influenced to some extent the topics discussed in the book."

I agree with these aims and judge that the author has met them very well indeed; hence, I recommend this excellent book for a one-semester, nonmathematical, introductory statistical course in the social sciences. It excels in the choice of exercises and illustrations drawn from important research publications in the social sciences, has many good examples and fine figures, tables, and charts to illustrate the important problems.

How does this book differ from some of its better competitors? First, it emphasizes, as the title states, statistical reasoning and inference rather than statistical techniques and manipulation. Rather than trying to provide a reference book on a large variety of statistical techniques, the author concentrates on a thorough presentation of the nature of statistical inference. He does this briefly in two chapters which contain the two most useful statistical methods for social scientists: Chapter 8, The Binomial Probability Model and Statistical Inference; and Chapter 9, Drawing Statistical Inferences from the Arithmetic Mean of a Large Sample.

Second, the statistical problems and examples are not alienated from their origins and their destinations. The problems and examples are established firmly in the substantive problems from which they arise, and in the problems of collection and processing which precede the data. The relation of sample to population is often and well developed. The meaning of statistical tools and of statistical inference is constantly emphasized; that statistics and probability statements are guides to action and to decisions is the spirit that pervades the presentation (although there is no formal presentation of statistical decision function theory). Chapter 2, The Components of Statistical Investigation, presents this approach early and well.

Third, the writing is rigorous and precise. The presentation is not mathematical and does not require a mathematical background, but it does demand close attention and careful reading. The style is clear, precise, and to the point, and will aid the student through the hard thinking that rigorous statistical inference demands.

Chapters 3 through 7 deal well with the necessary introductory topics of distributions, their locations and spread, and the elements of probability. Chapter 10 gives in 30 pages an excellent, lucid, and penetrating presentation of Elements of Sample Design. Each of the last two chapters presents an important technique for social scientists: Chapter 11, Chi-Square Procedures for Qualitative Data, and Chapter 12, The Linear Association of Two Quantitative Variables.

Some instructors will prefer to add one or two additional techniques to complete the course; for economists, perhaps time series and indexes; for psychologists, perhaps the difference of two correlated means and the elements of experimental design. The student will have to read his assignments thoroughly and sometimes repeatedly. The instructor will have to explain the finer points at greater length. But perhaps that is not too much to ask of a scientific subject in a scientific age. I think this is a good book.

LESLIE KISH

University of Michigan